



DESDE 1902  
INSTITUTO DE HIGIENE E  
MEDICINA TROPICAL  
UNIVERSIDADE NOVA DE LISBOA



UNIVERSIDADE  
**NOVA**  
DE LISBOA

# **EXPRESSÃO HETERÓLOGA DE ANTIGÉNIOS DO VÍRUS ZIKA EM *ESCHERICHIA COLI***

**CATARINA SOFIA FINO CASEIRO**

DISSERTAÇÃO PARA OBTENÇÃO DO  
GRAU DE MESTRE EM MICROBIOLOGIA MÉDICA

**DEZEMBRO, 2017**



DESDE 1902  
INSTITUTO DE HIGIENE E  
MEDICINA TROPICAL  
UNIVERSIDADE NOVA DE LISBOA



UNIVERSIDADE  
**NOVA**  
DE LISBOA

# **EXPRESSÃO HETERÓLOGA DE ANTIGÉNIOS DO VÍRUS ZIKA EM *ESCHERICHIA COLI***

**CATARINA SOFIA FINO CASEIRO**

DISSERTAÇÃO PARA OBTENÇÃO DO  
GRAU DE MESTRE EM MICROBIOLOGIA MÉDICA

Orientador: Professor Doutor Ricardo Parreira, Instituto de Higiene e Medicina Tropical  
Co-orientadora: Professora Doutora Aida Esteves, Instituto de Higiene e Medicina Tropical

**Unidade de Microbiologia Médica/Grupo de Virologia  
Instituto de Higiene e Medicina Tropical**

**DEZEMBRO, 2017**

## **AGRADECIMENTOS**

Agradeço especialmente ao meu orientador, Professor Doutor Ricardo Parreira, por todo o apoio dado no decorrer deste projeto. Ainda pela partilha de conhecimentos, dedicação, paciência, sentido crítico e disponibilidade demonstrada em todas as fases que levaram à concretização deste trabalho.

À minha co-orientadora, Professora Doutora Aida Esteves, pela sua atenção, conhecimentos transmitidos e ajuda dada, especialmente nos momentos iniciais deste trabalho.

Ao Professor Doutor Carlos São José, pela disponibilidade e conselhos na progressão deste projeto.

À Ângela Mendes, Técnica Superior na Unidade de Virologia (IHMT), pela simpatia, experiência e preciosa ajuda laboratorial.

Às colegas do IHMT, Andreia Lages, Filipa Geraldo Dias, Maria João Pinto da Costa, Patrícia Borges e Rute Marcelino, pela amizade, boa disposição e momentos de convívio criados ao longo desta etapa.

Aos meus amigos, em especial à Ana Carolina Varanda, Carolina Braga e Rúben Brandão, sobretudo pela amizade, apoio e motivação ao longo da minha vida.

À minha querida irmã Mariana, o meu grande orgulho.

E por último aos meus familiares, especialmente aos meus pais, pelo carinho e apoio incondicional, pois sem eles não seria possível chegar aqui. O meu sincero e sentido obrigado por acreditarem em mim.

*Dedico esta tese aos meus pais e irmã*

## RESUMO

O vírus Zika (ZIKV) é um arbovírus que pertence ao género *Flavivirus* da família *Flaviviridae*. Atualmente surge como um agente patogénico humano altamente relevante, não só pela possibilidade que demonstrou de se dispersar rapidamente, e de forma global, como pela sua associação ao desenvolvimento de uma série de malformações congénitas (incluindo microcefalia) e a síndrome de Guillain-Barré. Contudo, ainda pouco se conhece sobre a biologia do ZIKV, sendo para tal necessário o desenvolvimento de ferramentas básicas, nomeadamente, proteínas virais recombinantes ou anticorpos específicos, que permitam a utilização de uma grande variedade de protocolos experimentais.

Neste estudo realizou-se a otimização da produção/purificação de antígenos virais de ZIKV em *E. coli*, nomeadamente, da proteína estrutural da cápside (C) e da região C-terminal da proteína NS3, à qual está associada atividade enzimática de RNAhel (NS3hel). Para tal, as sequências virais de genótipo selvagem e versões sintéticas com codões otimizados foram clonadas no vetor pET-29a, de modo a possibilitar a sua expressão, em *E. coli*, como proteínas de fusão com caudas de hexa-histidina (His<sub>6</sub>) na extremidade C-terminal.

A expressão proteica heteróloga foi avaliada, quer em termos quantitativos, quer na solubilidade dos produtos obtidos. A identidade das proteínas C ( $\approx 13\text{kDa}$ ) e NS3hel ( $\approx 15\text{kDa}$ ) fundidas com caudas de His<sub>6</sub> (C-His<sub>6</sub> e NS3hel-His<sub>6</sub>, respetivamente) foi verificada através da sua purificação, em condições desnaturantes, por cromatografia de afinidade (em micro-escala) usando colunas com iões níquel imobilizados, e por *western blot*, utilizando anticorpos monoclonais anti-His<sub>6</sub>. As proteínas C-His<sub>6</sub> e NS3hel-His<sub>6</sub> foram, igualmente, sujeitas a purificação em condições nativas. Contudo, nestas condições, para além das proteínas em causa foi observada uma quantidade considerável de proteínas contaminantes. Ainda assim fica como prova de conceito o sucesso do processo de purificação em condições nativas.

**Palavras-chave:** Vírus Zika; proteína da cápside; domínio NS3hel; *Escherichia coli*; sistema pET.

## ABSTRACT

Zika virus (ZIKV) is an arbovirus classified in the *Flavivirus* genus of the *Flaviviridae* family. Presently, it appears as a highly relevant human pathogen, not only because of its ability to spread rapidly and globally, but also due to its association with the development of a series of congenital malformations (including microcephaly) and Guillain-Barré syndrome. However, the biology of ZIKV is still poorly understood, and its study requires the development of experimental basic tools, such as, recombinant viral proteins or specific antibodies, which will be used in a wide variety of experimental protocols.

In this study, we carried out the optimization of the production/purification of ZIKV viral antigens in *E. coli*, namely, the capsid structural protein (C) and the helicase domain of NS3 (NS3hel). For that purpose, wild type viral sequences and synthetic versions with optimized codons were cloned into the pET-29a vector, in order to enable their expression, in *E. coli*, as C-terminal hexa-histidine (His<sub>6</sub>) tagged fusion proteins.

Heterologous protein expression was assessed both in quantitative terms and the solubility of the products obtained. The identity of the C-terminal His<sub>6</sub> tagged fusion proteins C-His<sub>6</sub> ( $\approx$  13kDa) and NS3hel-His<sub>6</sub> ( $\approx$  15kDa) was verified by purification, under denaturing conditions, by immobilized metal affinity chromatography (micro-scale) and by western blot using anti-His<sub>6</sub> monoclonal antibodies. Both recombinant proteins were also subjected to purification under native conditions. Under these conditions, and despite their presence, a considerable amount of contaminating proteins were also co-purified. Nevertheless, the success of the purification process under native conditions remains as proof of concept.

**Keywords:** Zika virus; capsid protein; NS3hel domain; *Escherichia coli*; pET system.

## ÍNDICE GERAL

AGRADECIMENTOS .....	i
RESUMO.....	iii
ABSTRACT .....	iv
ÍNDICE GERAL .....	v
ÍNDICE DE FIGURAS .....	viii
ÍNDICE DE TABELAS .....	xiii
ABREVIATURAS .....	xiv
1. INTRODUÇÃO.....	1
1.1. Vírus Zika.....	1
1.1.1. ZIKV: o seu percurso histórico.....	2
1.1.2. Estrutura da partícula viral e organização genómica .....	4
1.1.3. Ciclo replicativo.....	6
1.1.4. Características das proteínas virais .....	7
1.1.4.1. Proteínas estruturais.....	7
1.1.4.2. Proteínas não-estruturais .....	8
1.1.5. Diversidade genética.....	10
1.1.6. Transmissão .....	10
1.1.6.1. Transmissão por mosquitos .....	10
1.1.6.2. Outros modos de transmissão de ZIKV .....	11
1.1.7. Patogénese e tropismo celular.....	12
1.1.8. Apresentação clínica e complicações da infeção por ZIKV .....	13
1.2. Expressão de proteínas recombinantes.....	15
1.2.1. Sistema de expressão bacteriano: <i>Escherichia coli</i> .....	16
1.2.1.1. Estratégias de otimização da produção de proteínas recombinantes em <i>E. coli</i> .....	18
1.3. Objetivos .....	21
2. MATERIAIS E MÉTODOS .....	23
2.1. Otimização das sequências nucleotídicas codificantes de C e NS3hel para ajustamento da preferência de utilização de codões em <i>E. coli</i> .....	23

2.2.	Extração de RNA total .....	23
2.3.	Síntese, por transcrição reversa, de DNA complementar, a partir de extratos de RNA total .....	24
2.4.	Delineamento de oligonucleótidos sintéticos iniciadores de reação em cadeia da polimerase .....	25
2.5.	Amplificação de sequências virais de ZIKV por PCR.....	27
2.5.1.	Amplificação de uma sequência viral para confirmação da presença de genoma de ZIKV .....	27
2.5.2.	Amplificação das regiões da C e NS3hel de sequências nucleotídicas de genótipo selvagem de ZIKV .....	28
2.5.3.	Amplificação das regiões codificantes da proteína C e do domínio NS3hel de ZIKV, a partir de sequências nucleotídicas sintéticas .....	28
2.6.	Análise por eletroforese em gel de agarose.....	29
2.7.	Purificação dos produtos de PCR.....	29
2.8.	Clonagem molecular das regiões que codificam C e NS3hel de ZIKV, a partir de sequências nucleotídicas sintéticas otimizadas e de genótipo selvagem.....	30
2.8.1.	Digestão enzimática do vetor de clonagem/expressão pET-29a e dos produtos de PCR.....	31
2.8.1.1.	Purificação dos produtos resultantes da digestão enzimática.....	31
2.8.2.	Ligação entre o vetor de clonagem/expressão pET-29a e produtos de PCR.....	32
2.8.3.	Preparação de células competentes de <i>E. coli</i> .....	32
2.8.4.	Transformação de <i>E. coli</i> por choque térmico .....	33
2.8.5.	Extração de DNA plasmídico pelo método de lise alcalina.....	33
2.8.6.	Análise de moléculas recombinantes de DNA plasmídico .....	34
2.8.7.	Sequenciação de DNA .....	35
2.9.	Expressão de proteínas recombinantes de ZIKV .....	35
2.9.1.	Teste de expressão proteica .....	36
2.9.2.	Testes de otimização de produção de proteínas recombinantes .....	37
2.9.3.	Avaliação da toxicidade proteica através da monitorização do crescimento bacteriano em meio líquido .....	37
2.9.4.	Testes de otimização de solubilidade proteica.....	37
2.9.5.	Avaliação do nível de expressão proteica por SDS-PAGE .....	39
2.9.6.	Recuperação da estrutura tridimensional de proteínas desnaturadas ( <i>Refolding</i> proteico).....	40



2.9.7.	Verificação da presença de cauda de His <sub>6</sub> , por purificação em condições desnaturantes, nas proteínas recombinantes produzidas .....	40
2.9.8.	Purificação de proteínas em condições nativas.....	41
2.9.9.	Verificação da presença de cauda de His <sub>6</sub> , por <i>western blot</i> , nas proteínas recombinantes produzidas .....	42
2.9.9.1.	Transferência de proteínas para uma membrana .....	42
2.9.9.2.	Bloqueio da membrana e incubação com anticorpos .....	42
2.9.9.3.	Deteção de proteínas por quimioluminescência .....	43
3.	RESULTADOS E DISCUSSÃO .....	44
3.1.	Otimização das sequências nucleotídicas codificantes de C e NS3hel para ajustamento da preferência de utilização de códons em <i>E. coli</i> .....	44
3.2.	Amplificação de sequências codificantes de ZIKV .....	47
3.2.1.	Confirmação, por RT-PCR, da presença de genoma de ZIKV numa suspensão viral.....	47
3.2.2.	Amplificação das regiões codificantes de C e NS3hel a partir de sequências de genótipo selvagem e versões sintéticas com sequências codificantes otimizadas.....	48
3.3.	Clonagem molecular das regiões que codificam C e NS3hel de ZIKV .....	51
3.4.	Expressão de proteínas recombinantes de ZIKV .....	54
3.4.1.	Testes de expressão proteica e de otimização de produção de proteínas recombinantes.....	55
3.4.2.	Testes de otimização de solubilidade proteica.....	64
3.4.3.	Recuperação da estrutura tridimensional de proteínas desnaturadas ( <i>refolding</i> proteico).....	75
3.4.4.	Purificação de proteínas em condições desnaturantes e nativas .....	76
3.4.5.	Verificação da presença de cauda de His <sub>6</sub> , por <i>western blot</i> , nas proteínas recombinantes produzidas .....	80
4.	CONCLUSÕES E PERSPETIVAS FUTURAS .....	84
5.	BIBLIOGRAFIA.....	86
6.	ANEXOS.....	97

## ÍNDICE DE FIGURAS

### 1. INTRODUÇÃO

**Figura 1.1.** Distribuição global da transmissão de ZIKV (mapa realizado a 29/8/2017; Adaptado de <https://ecdc.europa.eu/en/publications-data/current-zika-transmission-worldwide>).....4

**Figura 1.2.** Representação esquemática da partícula viral de ZIKV (Adaptado de [http://viralzone.expasy.org/6756?outline=all\\_by\\_protein](http://viralzone.expasy.org/6756?outline=all_by_protein)).....5

**Figura 1.3.** Representação esquemática da organização genômica e processamento da poliproteína para o género *Flavivirus*. A primeira caixa indica as regiões codificantes das componentes estruturais e não estruturais. As caixas abaixo indicam os precursores e proteínas maduras obtidas pela cascata de processamento proteolítico. As proteínas estruturais estão destacadas a violeta e as proteínas não estruturais (NS) a branco ou sobreadas de acordo com as subunidades enzimáticas. Os locais de clivagem estão também indicados, (♦) proteases celulares, (↓) protease viral, (▼) uma furina ou protease relacionada, e (?) proteases desconhecidas (Adaptado de Lindenbach *et al.*, 2013).....5

**Figura 1.4.** Representação simplificada do ciclo replicativo da família *Flaviviridae* (Adaptado de Lindenbach *et al.*, 2013).....7

**Figura 1.5.** Distribuição conhecida de *Aedes albopictus* na Europa a abril de 2017 (Adaptado de <https://ecdc.europa.eu/en/publications-data/aedes-albopictus-current-known-distribution-europe-april-2017>).....11

### 2. MATERIAIS E MÉTODOS

**Figura 2.1.** Esquema representativo do mapa do vetor pET-29a. O mapa apresenta a origem de replicação, o promotor natural e a sequência codificante de *lacI*, o local de clonagem múltipla (MCS) orientado a jusante do promotor/operador *T7lac* (seta a negrito) e um gene que confere resistência a canamicina (Kan).....30

**Figura 2.2.** Esquema representativo da região de clonagem/expressão do vetor pET-29a.....30

### 3. RESULTADOS E DISCUSSÃO

**Figura 3.1.** Confirmação da presença de genomas de ZIKV num extrato de RNA total, por amplificação por RT-PCR de um fragmento de  $\approx 500$ pb (após separação por eletroforese em gel de agarose a 1,5%). 1 - controlo negativo; 2 – produto de amplificação desejado (ver texto); M – marcador de massa molecular (*NZYDNA Ladder VI*).....47

**Figura 3.2.** Amplificação por PCR das regiões codificantes de C e NS3hel de sequências nucleotídicas de genótipo selvagem (**A**) e versões sintéticas com codões otimizados (**B**) e correspondente à região codificante de C de sequência nucleotídica sintética otimizada com uma deleção na região 3' (codificando um segmento hidrofóbico) (**C**), após separação por eletroforese em gel de agarose a 1,5%. M-*NZYDNA Ladder VI*. Controlos negativos: **A** - pistas 3 e 4, **B** - pistas 1 e 3, e **C** -pista 1. **A** - 1 e 2 indicam os fragmentos de DNA correspondentes às regiões codificantes de C e NS3hel, respetivamente. **B**- 2 e 4 indicam os fragmentos de DNA correspondentes às regiões codificantes de C e NS3hel, respetivamente. **C**- 2 indica o fragmento de DNA correspondente à região codificante de C.....51

**Figura 3.3.** Rastreio por PCR de derivados de pET-29a contendo sequências de ZIKV, após separação por eletroforese em gel de agarose a 1,5%. M - *NZYDNA Ladder VI*. As caixas retangulares indicam os clones bacterianos onde foi detetada a presença de plasmídeos recombinantes. (**A**) 1 e (**B**) 10 – controlos positivos; (**A**) 2 e (**B**) 11 – controlos negativos; (**A**) e (**B**) - amplificação, por PCR, dos insertos correspondentes às regiões codificantes de C (3 a 8) e NS3hel (12 a 17) nas suas versões sintéticas com codões otimizados.....54

**Figura 3.4.** Observação do efeito de toxicidade sobre culturas de *E. coli BL21(DE3)pLysS* e *Rosetta(DE3)pLysS* transformadas com *pET-29a::C(ot)*, quando induzida a expressão da proteína C, a 37°C durante 3h. 1 e 2 – culturas induzida (I) e não induzida (N.I) de *E. coli BL21(DE3)pLysS*, respetivamente. 3 e 4 - culturas I e N.I de *E. coli Rosetta(DE3)pLysS*, respetivamente.....56

**Figura 3.5.** Análise de extratos de proteínas solúveis totais de culturas N.I e I de *E. coli BL21(DE3)pLysS* (B) e *Rosetta(DE3)pLysS* (R), expressando NS3hel(wt) (pistas 1, 2, 5 e 6), NS3hel(ot) (pistas 3, 4, 7 e 8), C(wt) (pistas 10-13; 18-21) e C(ot) (pistas 14-17; 22-29), em géis de tricina-SDS-PAGE (12%). Os tempos de indução e as temperaturas a

que estas ocorreram estão indicados na figura. M-marcador de proteínas com massas moleculares conhecidas (*Precision Plus Protein™ Dual Color Standards*).....59

**Figura 3.6.** Curvas de crescimento de culturas controlo (ausência de indução da expressão proteica heteróloga, indicadas por (-)IPTG) ou de culturas em que a esta foi induzida pela adição de IPTG [indicadas por (+)IPTG]. A indução da expressão proteica pretendida foi iniciada a D.O.<sub>600nm</sub>, de  $\approx 0,4$ , seguida de incubação a 37°C durante 3h....61

**Figura 3.7.** Análise, após separação por tricina-SDS-PAGE a 12%, de extratos proteicos totais derivados de culturas N.I e I de *E. coli BL21(DE3)pLysS* (B: 1, 3-6) e *Rosetta(DE3)pLysS* (R: 2, 7-10), expressando C(ot-del), a 37°C durante 3h, para avaliação do parâmetro, heterogeneidade da população bacteriana. M - marcador de proteínas com massas moleculares conhecidas (*Precision Plus Protein™ Dual Color Standards*).....63

**Figura 3.8.** Análise, após separação por tricina-SDS-PAGE a 12%, de extratos proteicos totais derivados de culturas N.I (9) e I de *E. coli Rosetta(DE3)pLysS* (R) expressando NS3hel(ot) (5-9) ou C(ot-del) (1-4), a 37°C durante 1h, 2h, 3h e 5h, para avaliação do parâmetro, ensaio de cinética de produção.....64

**Figura 3.9.** Análise, após separação por SDS-PAGE a 13%, de um extrato de proteínas solúveis totais de uma cultura controlo (ausência de indução da expressão proteica heteróloga; 1) e de extratos proteicos correspondendo às frações solúvel (FS) e insolúvel (FI), de culturas onde foi induzida a produção heteróloga de NS3hel (2, 3, 6 e 7) ou C (4, 5, 8 e 9), a 37°C durante 1h e 3h, para verificação da solubilidade proteica....66

**Figura 3.10.** Análise, após separação por SDS-PAGE a 13%, **A)** de um extrato de proteínas solúveis totais de uma cultura controlo (1) e de extratos proteicos correspondendo às frações solúvel (FS) e insolúvel (FI), de culturas onde foi induzida a produção heteróloga de NS3hel (2, 3, 6, 7, 10 e 11) ou C (4, 5, 8, 9, 12 e 13). **B)** de extratos de proteínas solúveis totais de culturas controlo (1 e 10) e de uma cultura onde foi induzida a produção heteróloga de NS3hel (19), e de extratos proteicos correspondendo às frações solúvel e insolúvel, de culturas onde foi induzida a produção heteróloga de NS3hel (2, 3, 6, 7, 11, 12, 15 e 16) ou C (4, 5, 8, 9, 13, 14, 17 e 18). Os tempos de indução e as temperaturas a que estas ocorreram estão indicados na figura.....69

**Figura 3.11.** Análise, após separação por SDS-PAGE a 13%, de extratos de proteínas solúveis totais de culturas controlo (1 e 10) e de uma cultura onde foi induzida a produção

heteróloga de NS3hel (19), e de extratos proteicos correspondendo às frações solúvel (FS) e insolúvel (FI), de culturas onde foi induzida a produção heteróloga de NS3hel (2, 3, 6, 7, 11, 12, 15 e 16) ou C (4, 5, 8, 9, 13, 14, 17 e 18). A indução da expressão proteica pretendida foi iniciada a D.O.<sub>600nm</sub>, de  $\approx 0,4$ , 0,8 ou 1,2 seguida de incubação a 37°C durante 1h, 2h ou 3h.....71

**Figura 3.12.** Análise, após separação por SDS-PAGE a 13% **A)** de um extrato de proteínas solúveis totais de uma cultura controlo (1) e de extratos proteicos correspondendo às frações solúvel (FS) e insolúvel (FI), de culturas onde foi induzida a produção heteróloga de NS3hel (2, 3, 6 e 7) ou C (4, 5, 8 e 9). **B)** de extratos de proteínas solúveis totais de culturas controlo (1 e 11) e de culturas onde foi induzida a produção heteróloga de NS3hel (10 e 20), e de extratos proteicos correspondendo às frações solúvel e insolúvel, de culturas onde foi induzida a produção heteróloga de NS3hel (2, 3, 6, 7, 12, 13, 16 e 17) ou C (4, 5, 8, 9, 14, 15, 18 e 19). As concentrações de D-sorbitol testadas, os tempos de indução e a temperatura a que as induções ocorreram estão indicados na figura.....72

**Figura 3.13.** Análise, após separação por SDS-PAGE a 13% **A)** de extratos de proteínas solúveis totais de culturas controlo (1 e 10) e de uma cultura onde foi induzida a produção heteróloga de NS3hel (19) e de extratos proteicos correspondendo às frações solúvel (FS) e insolúvel (FI), de culturas onde foi induzida a produção heteróloga de NS3hel (2, 3, 6, 7, 11, 12, 15 e 16) ou C (4, 5, 8, 9, 13, 14, 17 e 18). **B)** de extratos de proteínas solúveis totais derivados de uma cultura controlo (1) e de uma onde foi induzida a produção heteróloga de NS3hel (10), e de extratos proteicos, obtidos sem recurso a sonicação (NS), correspondendo às frações solúvel e insolúvel, de culturas onde foi induzida a produção heteróloga de NS3hel (2, 3, 6 e 7) ou C (4, 5, 8 e 9 ). As concentrações de D-sorbitol testadas, os tempos de indução e a temperatura a que as induções ocorreram estão indicados na figura.....74

**Figura 3.14.** **A)** Observação de um extrato proteico total depois do processo de diálise e o qual é derivado de uma cultura onde foi induzida a produção heteróloga de NS3hel. O aspeto “leitoso” da suspensão proteica é indicativo da precipitação de proteínas. **B)** Análise, após separação por SDS-PAGE a 13%, de um extrato proteico total (5) e de extratos proteicos correspondendo às frações solúvel (FS) e insolúvel (FI), antes (AD) e depois (DD) do processo de diálise, os quais são derivados de culturas onde foi induzida

a produção heteróloga de NS3hel, a 37°C durante 3h.....76

**Figura 3.15.** Análise, após separação por SDS-PAGE a 13% dos produtos resultantes da purificação por cromatografia de afinidade com iões níquel das proteínas NS3hel (**A**) e C (**B**) em condições desnaturantes. **A** e **B** - 1 e 2 – extratos proteicos de culturas controlo e correspondente à fração solúvel (FS) de culturas onde foi induzida a produção heteróloga, respetivamente. Filtrados correspondentes aos passos de fixação (3) e de lavagens (4 e 5). 6 e 7 – eluatos. 8 e 9 – extrato de proteína solúvel total (E.P.T) e correspondente à fração insolúvel (FI), de culturas onde foi induzida a produção heteróloga, respetivamente. O tempo de indução e a temperatura a que esta ocorreu estão indicados na figura.....78

**Figura 3.16.** Análise, após separação por SDS-PAGE a 13%, dos produtos resultantes da purificação por cromatografia de afinidade com iões níquel das proteínas NS3hel (**A**) e C (**B**) em condições nativas. **A** e **B** - 1 e 2 – extratos proteicos de culturas controlo e correspondente à fração solúvel (FS) de culturas onde foi induzida a produção heteróloga, respetivamente. Filtrados correspondentes aos passos de fixação (3) e de lavagens (4 e 5). 6 e 7 – eluatos. 8 e 9 – extrato proteico correspondente à fração insolúvel (FI) e extrato de proteína solúvel total (E.P.T) de culturas onde foi induzida a produção heteróloga, respetivamente. O tempo de indução e a temperatura a que esta ocorreu estão indicados na figura.....80

**Figura 3.17. A)** Análise, após separação por SDS-PAGE a 13%, de extratos proteicos totais de culturas induzidas, com (7 e 8) e sem (1 e 2) suplementação do meio LB líquido com 0,5M D-sorbitol, a 37°C durante 3h, de eluatos resultantes do processo de purificação em condições desnaturantes (3 e 4) e de extratos proteicos totais derivados de culturas controlo (9 e 10). **B)** Detecção da presença de caudas de His<sub>6</sub> nas proteínas NS3hel e C por *western blot* (pistas equivalentes em A). 5 - M - marcador de proteínas com massas moleculares conhecidas (*Precision Plus Protein™ Dual Color Standards*).....81

## ÍNDICE DE TABELAS

### 2. MATERIAIS E MÉTODOS

<b>Tabela 2.1.</b> Condições utilizadas na reação de RT.....	24
<b>Tabela 2.2.</b> Informação relativa ao conjunto de <i>primers</i> utilizados nas amplificações, por PCR, de uma sequência genômica (coordenadas 165-661 na sequência de genoma completo, LC002520) para detecção de ZIKV, e das regiões da C e NS3hel de genótipo selvagem.....	26
<b>Tabela 2.3.</b> Informação relativa ao conjunto de <i>primers</i> utilizados nas amplificações, por PCR, das sequências nucleotídicas sintéticas otimizadas que codificam C e NS3hel, e ao <i>primer R</i> redesenhado para a amplificação da região C.....	26
<b>Tabela 2.4.</b> Condições de reação utilizadas na amplificação, por PCR, para confirmação da presença de genoma de ZIKV.....	27
<b>Tabela 2.5.</b> Condições de reação utilizadas na amplificação, por PCR, das regiões codificantes de C e NS3hel, a partir de sequências nucleotídicas de genótipo selvagem e versões sintéticas com códons otimizados.....	29
<b>Tabela 2.6.</b> Condições de reação utilizadas na amplificação, por PCR, das regiões codificantes de C e NS3hel para análise de moléculas recombinantes de DNA plasmídico.....	35

### 3. RESULTADOS E DISCUSSÃO

<b>Tabela 3.1.</b> Frequência absoluta de códons raros em <i>E. coli</i> nas regiões codificantes dos antígenos C e NS3hel, na sequência correspondente ao genoma completo de ZIKV (número de acesso LC002520).....	46
---	----

## ABREVIATURAS

<b>ADB</b>	Tampão de dissolução de agarose
<b>Amp</b>	Ampicilina
<b>APS</b>	Persulfato de amónio
<b>Arbovírus</b>	Do inglês <u>arthropode-borne virus</u>
<b>Arg</b>	Arginina
<b>C</b>	Proteína estrutural da cápside
<b>°C</b>	Graus Celsius
<b>Cam</b>	Cloranfenicol
<b>cDNA</b>	DNA complementar, do inglês <u>Complementary Deoxyribonucleic Acid</u>
<b>CHIKV</b>	Vírus Chikungunya, do inglês <u>Chikungunya Virus</u>
<b>CHO</b>	Linha celular derivada de ovário de hamster chinês
<b>CI</b>	Corpos de inclusão
<b>D.O.</b>	Densidade ótica
<b>DENV</b>	Vírus da Dengue, do inglês <u>Dengue Virus</u>
<b>DMSO</b>	Dimetilsulfóxido
<b>DNA</b>	Ácido desoxirribonucleico, do inglês <u>Deoxyribonucleic Acid</u>
<b>DNase</b>	<u>Desoxirribonuclease</u>
<b>dNTPs</b>	Desoxirribonucleótidos 5'-trifosfatados, do inglês <u>Deoxynucleotide Triphosphates</u>
<b>E</b>	Glicoproteína do invólucro
<b><i>E. coli</i></b>	<i>Escherichia coli</i>
<b>EDTA</b>	Etilenodiaminotetracetato
<b>E.U.A</b>	<u>Estados Unidos da América</u>
<b>Fh8</b>	Proteína de ligação ao cálcio de 8kDa
<b>g</b>	Aceleração da gravidade
<b>Gly</b>	Glicina
<b>GST</b>	Glutathione S-transferase
<b>h</b>	Hora
<b>HEK293</b>	Linha celular derivada de rim embrionário humano
<b>Hi5</b>	Linha celular de inseto, <i>Trichoplusia ni</i>



<b>His<sub>6</sub></b>	Cauda de hexa-histidina
<b>hNPC</b>	Célula progenitora neuronal humana
<b>I</b>	Induzida
<b>Ile</b>	Isoleucina
<b>IPTG</b>	Isopropil-β-D-tiogalactopiranosídeo
<b>ISF</b>	Flavivírus específicos de insetos, do inglês <i>Insect-Specific Flavivirus</i>
<b>JEV</b>	Vírus da encefalite Japonesa, do inglês <i>Japanese Encephalitis Virus</i>
<b>Kan</b>	Canamicina
<b>KCl</b>	Cloreto de potássio
<b>kDa</b>	Kilo(Daltons) = 10 <sup>3</sup> Daltons
<b>LB</b>	Meio de cultura <u>L</u> uria- <u>B</u> ertani
<b>Leu</b>	Leucina
<b>M</b>	Glicoproteína de membrana viral
<b>m</b>	Mili, prefixo que designa 10 <sup>-3</sup>
<b>M</b>	Molar
<b>MBP</b>	Proteína de ligação a maltose
<b>MCS</b>	Local de clonagem múltipla, do inglês <i>Multiple Cloning Site</i>
<b>MgCl<sub>2</sub></b>	Cloreto de magnésio
<b>MgSO<sub>4</sub></b>	Sulfato de magnésio
<b>min</b>	Minuto
<b>ml</b>	Mililitro = 10 <sup>-3</sup> litros
<b>mM</b>	Milimolar
<b>mRNA</b>	Ácido ribonucleico mensageiro, do inglês <i>Messenger Ribonucleic Acid</i>
<b>N.I</b>	Não induzida
<b>Na<sub>2</sub>HPO<sub>4</sub></b>	Fosfato dissódico
<b>NaCl</b>	Cloreto de sódio
<b>NaH<sub>2</sub>PO<sub>4</sub>.H<sub>2</sub>O</b>	Dihidrogenofosfato de sódio monohidratado
<b>NaOH</b>	Hidróxido de sódio
<b>NCBI</b>	Do inglês <i>National Center for Biotechnology Information</i>
<b>ng</b>	Nanógrama =10 <sup>-9</sup> gramas

<b>nm</b>	Nanómetro = $10^{-9}$ metros
<b>NS3hel</b>	Proteína não estrutural 3 (domínio ao qual é atribuída atividade de helicase de RNA)
<b>NS5MTase</b>	Proteína não estrutural 5 (domínio ao qual é atribuída atividade de metiltransferase)
<b>nt</b>	Nucleótido
<b>NTPase</b>	Nucleósido trifosfatase
<b>NusA</b>	Proteína de terminação/anti-terminação
<b>OMS</b>	Organização Mundial de Saúde
<b>ON</b>	Durante a noite, do inglês <i>overnight</i>
<b>ORF</b>	Grelha de leitura aberta, do inglês <i>Open Reading Frame</i>
<b>pb</b>	Pares de bases
<b>PCR</b>	Reação em cadeia da polimerase, do inglês <i>Polymerase Chain Reaction</i>
<b>PEG</b>	Polietilenoglicol
<b>PF</b>	Polinésia Francesa
<b>pH</b>	Simétrico do logaritmo decimal da concentração hidrogeniônica de uma solução
<b>poli-A</b>	Cauda poliadenilada
<b>primer F</b>	<i>Primer</i> direto
<b>primer R</b>	<i>Primer</i> reverso
<b>prM</b>	Glicoproteína precursora de membrana
<b>Pro</b>	Prolina
<b>PVDF</b>	Fluoreto de polivinilideno
<b>rbs</b>	Local de ligação do ribossoma, do inglês <i>ribosome-binding site</i>
<b>RE</b>	Retículo endoplasmático
<b>RNA</b>	Ácido ribonucleico, do inglês <i>Ribonucleic Acid</i>
<b>RNAhel</b>	Helicase de RNA
<b>RNase</b>	<i>Ribonuclease</i>
<b>rpm</b>	Rotações por minuto
<b>RpRd</b>	Polimerase de RNA dependente de RNA, do inglês <i>RNA-dependent RNA polymerase</i>

<b>RT</b>	Transcrição reversa, do inglês <i>Reverse Transcription</i>
<b>RTPase</b>	Trifosfatase de RNA
<b>RT-PCR</b>	Reação em cadeia da polimerase antecedida de transcrição reversa
<b>SDS</b>	Dodecilsulfato de sódio, do inglês <i>Sodium Dodecyl Sulphate</i>
<b>SDS-PAGE</b>	Eletroforese em gel de poliacrilamida na presença de dodecilsulfato de sódio
<b>seg</b>	Segundos
<b>Sf9</b>	Linha celular de inseto, <i>Spodoptera frugiperda</i>
<b>SGB</b>	Síndrome de Guillain-Barré
<b>SUMO</b>	Pequena proteína modificadora semelhante a ubiquitina
<b>TAE</b>	Tampão Tris-acetato EDTA
<b>TE</b>	Tampão Tris-EDTA
<b>TEMED</b>	N, N, N', N'- tetrametiletilenodiamina
<b>TF</b>	Do inglês <i>trigger factor</i>
<b>Tm</b>	Temperatura de dissociação, do inglês <i>melting temperature</i>
<b>Tris</b>	Tris(hidroximetil)aminometano
<b>Tris-HCl</b>	Tris Hidrocloreto
<b>tRNA</b>	Ácido ribonucleico de transferência, do inglês <i>Transfer Ribonucleic Acid</i>
<b>Trx</b>	Tiorredoxina
<b>TSS</b>	Solução de transformação e armazenamento, do inglês <i>Transformation and Storage Solution</i>
<b>U</b>	Unidade
<b>UTR</b>	Região não traduzida, do inglês, <i>Untranslated Region</i>
<b>UV</b>	Ultravioleta
<b>V</b>	Volt
<b>WNV</b>	Vírus do Nilo Ocidental, do inglês <i>West Nile Virus</i>
<b>YFV</b>	Vírus da febre-amarela, do inglês <i>Yellow Fever Virus</i>
<b>ZIKV</b>	Vírus Zika, do inglês <i>Zika Virus</i>
<b>µg</b>	Microgramas = 10 <sup>-6</sup> gramas
<b>µl</b>	Microlitros = 10 <sup>-6</sup> litros
<b>pmol</b>	Picomole = 10 <sup>-12</sup> moles

## 1. INTRODUÇÃO

### 1.1. Vírus Zika

O vírus Zika (ZIKV) é um arbovírus (do inglês *arthropode-borne virus*) que pertence ao género *Flavivirus* da família *Flaviviridae* (Musso e Gubler, 2016). O termo arbovírus surge como uma definição ecológica, e não taxonómica, designando vírus que permanecem na natureza essencialmente por transmissão biológica, isto é, infectando hospedeiros vertebrados suscetíveis aos quais são transmitidos pela picada de artrópodes hematófagos (designados de vetores) por eles igualmente infectados (WHO, 1967). Atualmente, o género *Flavivirus* inclui 53 espécies virais (Simmonds *et al.*, 2017), das quais 27 são transmitidas por mosquitos, 12 por carraças e 14 são flavivírus zoonóticos sem vetor artrópode conhecido (Blitvich e Firth, 2017). Um quarto grupo de vírus encontrados unicamente em insetos, designados flavivírus específicos de insetos [do inglês *insect-specific flavivirus* (ISF)], possivelmente também será colocado neste género (Roundy *et al.*, 2017). Os principais vetores artrópodes de arbovírus patogénicos para o Homem são os mosquitos, e dentro do género *Flavivirus*, ZIKV é transmitido por esse grupo de vetores, em particular por membros do género *Aedes*, sendo *Aedes aegypti* o seu principal vetor. ZIKV é filogeneticamente próximo a flavivírus com importância na saúde pública internacional, e dos quais são exemplo o vírus da encefalite Japonesa (JEV), o vírus do Nilo Ocidental (WNV), o vírus da Dengue (DENV) e o vírus da febre-amarela (YFV) (Song *et al.*, 2017).

Atualmente ZIKV surge como um arbovírus emergente com potencial quasi-pandémico, o qual foi reconhecido aquando da sua dispersão epidémica a partir do Brasil (iniciada em 2015), e que rapidamente se alastrou por toda a América do Sul e Central, chegando aos Estados Unidos da América (E.U.A) (Bardina *et al.*, 2017; Baud *et al.*, 2017). Até então ZIKV não era considerado um agente patogénico humano relevante. No entanto, os surtos generalizados que se estenderam por todo o continente americano (Bardina *et al.*, 2017) e a associação de infeção por ZIKV quer com uma série de síndromes congénitas, quer com a síndrome de Guillain-Barré (SGB) (Barzon *et al.*, 2016b) levaram a Organização Mundial de Saúde (OMS) a declarar ZIKV, como uma emergência de saúde pública de interesse internacional entre fevereiro e novembro de 2016 (WHO, 2016a; Barzon *et al.*, 2016b; WHO, 2016b). Até ao momento não se

encontra disponível nenhuma vacina ou composto antiviral para prevenção ou tratamento específico da infecção pelo ZIKV. O aumento do número de casos e o potencial risco de disseminação global tornam ZIKV um grande desafio para a saúde pública mundial (Lazear e Diamond, 2016).

### **1.1.1. ZIKV: o seu percurso histórico**

ZIKV foi isolado em 1947, a partir de uma amostra de soro de um macaco *Rhesus* (*Macaca mulatta*) febril, servindo de sentinela no estudo de vigilância de YFV na floresta de Zika no Uganda. Mais tarde em 1948 foi isolado em mosquitos, *Aedes africanus*, na mesma região (Dick *et al.*, 1952). O primeiro caso confirmado de infecção humana por ZIKV parece ter sido o reportado por Simpson entre 1962 e 1963 (Simpson, 1964; Wikan e Smith, 2016). A realização de inquéritos sorológicos no decorrer do século, em conjunto com a evidência virológica, permitiram sugerir a endemicidade de ZIKV em África atingindo países como o Senegal, Serra Leoa, Nigéria, Gabão, República Centro-Africana, e Uganda, e na Ásia, circulando na Índia, Tailândia, Malásia, Vietname, Indonésia e Filipinas (Song *et al.*, 2017). É sempre necessária uma interpretação cuidadosa de dados epidemiológicos que assentam, em grande parte, na análise de dados sorológicos, devido a uma extensiva reatividade cruzada entre os membros do género *Flavivirus*. Contudo, e em grande parte, mais recentemente estes mesmos dados foram confirmados pelo isolamento de ZIKV em vetores e hospedeiros vertebrados na maioria destes países (Musso e Gubler, 2016; Weaver *et al.*, 2016). A ausência de quadros clínicos severos e de grandes surtos, e o subdiagnóstico da infecção permitiu a transmissão silenciosa de ZIKV até 2007, com apenas 14 infeções confirmadas em humanos até essa data (Musso e Gubler, 2016).

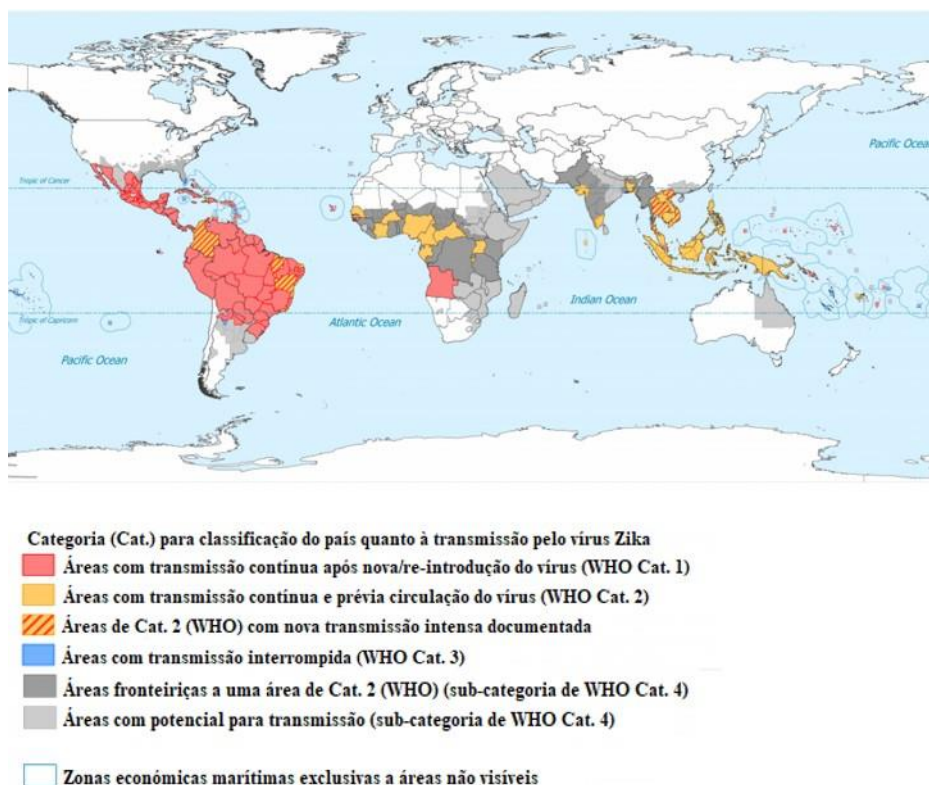
No ano de 2007, na Federação dos Estados da Micronésia, e em particular na ilha de Yap, ocorreu a primeira epidemia atribuída a ZIKV, constituindo o primeiro foco fora de África e da Ásia, com uma estimativa final apontando para que, aproximadamente 73% da população com idade  $\geq 3$  anos tenha sido infetada (Duffy *et al.*, 2009; Musso e Gubler, 2016). Nos anos seguintes apenas foram reportados casos esporádicos de infecção pelo ZIKV no Sudeste Asiático, em países como a Tailândia, o Camboja, a Malásia, a Indonésia e as Filipinas (Song *et al.*, 2017).

Em outubro de 2013, o vírus reemergiu na Polinésia Francesa (PF) provocando a segunda grande epidemia atingindo a população humana. Após a epidemia inquéritos sorológicos realizados sugerem um valor de seroprevalência de  $\approx 50\%$ , com a maioria dos indivíduos infectados a não procurarem ajuda médica, por apresentarem sintomas leves de febre Zika ou infecção assintomática (Musso *et al.*, 2014b; Aubry *et al.*, 2017). O vírus durante e após a epidemia na PF disseminou-se rapidamente pelas ilhas do Pacífico, provocando pequenos surtos, em 2014, na Nova Caledônia, Ilhas Cook e Ilha de Páscoa e, em 2015, em Vanuatu, Ilhas Salomão, Samoa e Fiji. (Musso *et al.*, 2014b; Musso e Gubler, 2016).

No início de 2015, ZIKV emerge, pela primeira vez, no continente americano, no nordeste do Brasil (estado da Bahia) (Zanluca *et al.*, 2015) em associação com um surto de doença exantemática aguda (Cardoso *et al.*, 2015). No fim de 2015, a transmissão autóctone do vírus foi confirmada em 18 estados brasileiros (PAHO, 2015) e estimado um número entre 440.000 a 1.300.000 casos suspeitos de infecção pelo ZIKV (Hennessey *et al.*, 2016), sendo que a maioria foram reportados nas regiões nordeste (Bahia) e sudeste (Rio de Janeiro) do país (Brasil *et al.*, 2016; Brito e Cordeiro, 2016; Faria *et al.*, 2017). Segundo análises filogenéticas, a introdução de uma linhagem genotípica asiática no continente americano deverá ter ocorrido por volta do final de 2013, ou seja, pelo menos um ano antes da sua detecção. Também foi possível verificar a partilha de um antepassado comum com a estirpe circulante na PF (Faria *et al.*, 2016).

No início de 2017, a maioria dos países da América do Sul e Central e Caraíbas haviam relatado circulação ativa do vírus (WHO, 2017), sendo inferido a origem das linhagens de ZIKV, a região nordeste do Brasil (Faria *et al.*, 2017).

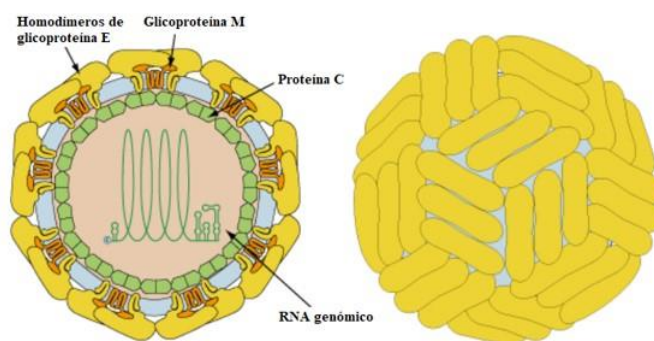
De uma forma genérica, têm sido registados, um pouco por todo o lado, um crescimento no número de casos importados de infeções a ZIKV (viajantes infectados de retorno ao país não endémico), provocando o aumento do risco de disseminação deste vírus a áreas nas quais foi já registada a presença de potenciais vetores competentes para a sua transmissão (Musso e Gubler, 2016). Os casos foram reportados em países como a Alemanha, Austrália, Espanha, Canadá, França, Holanda e Portugal (Song *et al.*, 2017).



**Figura 1.1.** Distribuição global da transmissão de ZIKV (mapa realizado a 29/8/2017; Adaptado de <https://ecdc.europa.eu/en/publications-data/current-zika-transmission-worldwide>).

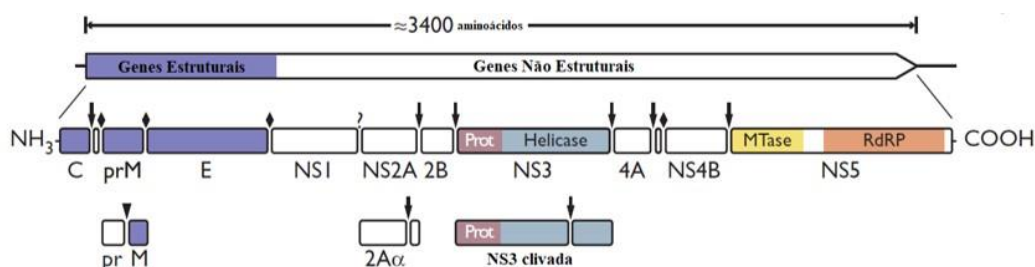
### 1.1.2. Estrutura da partícula viral e organização genómica

Tal como foi referido acima, ZIKV é um dos membros da família *Flaviviridae* e apresenta um virião de  $\approx 50\text{nm}$  de diâmetro, limitado por um invólucro lipídico, onde estão inseridas 180 cópias da glicoproteína do invólucro (E) e da glicoproteína de membrana viral [M, proteína resultante da clivagem da glicoproteína precursora de membrana (prM)], organizadas em simetria icosaédrica (Sirohi *et al.*, 2016; Kostyuchenko *et al.*, 2016; Song *et al.*, 2017). Nos viriões maduros, as proteínas E encontram-se estruturadas em homodímeros, consistindo a partícula viral em 90 dímeros de proteína E, dispostos antiparalelamente (Perera *et al.*, 2008; Saiz *et al.*, 2016). M é uma proteína de pequenas dimensões, encontrando-se sobreposta pela E (Kostyuchenko *et al.*, 2016). O invólucro reveste a nucleocápside de  $\approx 30\text{nm}$ , a qual é constituída por múltiplas cópias da proteína estrutural da cápside (C) associada ao genoma viral (Barzon *et al.*, 2016b).



**Figura 1.2.** Representação esquemática da partícula viral de ZIKV (Adaptado de [http://viralzone.expasy.org/6756?outline=all\\_by\\_protein](http://viralzone.expasy.org/6756?outline=all_by_protein)).

O genoma viral consiste numa única cadeia linear de RNA de polaridade positiva com  $\approx 10.800$  nucleótidos (nt). O genoma encontra-se estabilizado por uma estrutura *cap* na extremidade 5' e não apresenta uma cauda poliadenilada (poli-A) na extremidade 3'. A única grelha de leitura aberta (ORF, do inglês *Open Reading Frame*) que apresenta encontra-se flanqueada por regiões não traduzidas (UTRs, do inglês *Untranslated Regions*), localizadas em ambas as extremidades do genoma, e por isso designadas 5'UTR (com cerca de  $\approx 100$ -nt) e 3'UTR (com cerca de  $\approx 400$ -nt), respetivamente. A região 5'UTR apresenta estruturas secundárias semelhantes às que aparentemente caracterizam os genomas de diferentes flavivírus, com funções na regulação da tradução da poliproteína viral. A região 3'UTR apresenta estruturas e sequências conservadas entre flavivírus, as quais estão implicadas na tradução e replicação do genoma viral. A tradução da ORF produz uma poliproteína com  $\approx 3400$  resíduos, a qual é, subsequentemente, clivada pela atividade combinada de proteases virais e celulares, resultando na formação de 10 proteínas. As 3 proteínas estruturais N-terminal são C, prM e E, com funções na formação dos viriões e as 7 proteínas não estruturais C-terminal são NS1, NS2A/B, NS3, NS4A/B e NS5, com funções regulatórias e enzimáticas (Lindenbach *et al.*, 2007; Yun *et al.*, 2016; Barzon *et al.*, 2016b).



**Figura 1.3.** Representação esquemática da organização genómica e processamento da poliproteína para o género *Flavivirus*. A primeira caixa indica as regiões codificantes das componentes estruturais e não



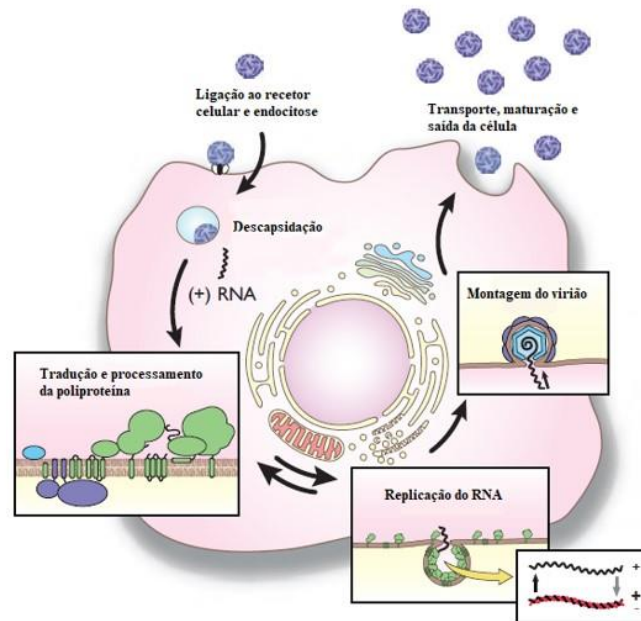
estruturais. As caixas abaixo indicam os precursores e proteínas maduras obtidas pela cascata de processamento proteolítico. As proteínas estruturais estão destacadas a violeta e as proteínas não estruturais (NS) a branco ou sobreadas de acordo com as subunidades enzimáticas. Os locais de clivagem estão também indicados, (◆) proteases celulares, (↓) protease viral, (▼) uma furina ou protease relacionada, e (?) proteases desconhecidas (Adaptado de Lindenbach *et al.*, 2013).

### 1.1.3. Ciclo replicativo

Os dados disponíveis sobre o ciclo replicativo de ZIKV são escassos, contudo, presume-se que a sua estratégia replicativa seja semelhante à dos restantes flavivírus (esquemáticamente representado na Figura 1.4). A replicação viral ocorre no citoplasma das células infetadas (Barzon *et al.*, 2016b).

Os flavivírus iniciam o ciclo replicativo com a ligação da partícula viral à superfície da célula hospedeira através da interação entre as glicoproteínas E e um ou mais recetores celulares. De momento, para ZIKV, o recetor celular não se encontra identificado, no entanto, estudos *in vitro* sugerem como candidatos, um recetor de lectina tipo C, DC-SIGN, dois membros da família proteica TAM, AXL e Tyro3, que são recetores transmembranares tirosina cinase e a glicoproteína TIM-1 (Hamel *et al.*, 2015). As partículas virais são internalizadas por endocitose. Nos endossomas, o abaixamento do pH resulta no rearranjo estrutural da glicoproteína E, o que promove a fusão do invólucro viral com a membrana endossomal, levando à descapsidação do virião e a libertação do RNA viral no citoplasma (Lindenbach *et al.*, 2007; Kaufmann e Rossmann, 2011; Song *et al.*, 2017). O RNA genómico viral é diretamente traduzido, formando uma poliproteína, que de seguida é processada por proteases. Para a replicação viral são formados complexos replicativos que se associam a membranas intracelulares (muito provavelmente derivadas do retículo endoplasmático-RE), em associação às quais ocorre a replicação do genoma. Este processo começa com a síntese de um RNA intermediário de polaridade negativa, servindo como molde para a síntese de novas cadeias de RNA de polaridade positiva. Estas servirão como RNA mensageiro (mRNA) e RNA genómico, o qual será encapsidado nos novos viriões em formação. O RNA viral é encapsidado pela proteína C formando a nucleocápside que, de seguida, é envolvida pela membrana do RE juntamente com as glicoproteínas prM e E, e originando partículas virais imaturas. Estas são transportadas para a região distal do aparelho de Golgi (*trans-Golgi*) onde sofrem maturação, a qual envolve a clivagem de prM (dando origem a proteínas M maduras), e a glicosilação de E. Os viriões maduros eventualmente abandonam as células hospedeiras

por exocitose (Perera *et al.*, 2008; Sampath e Padmanabhan, 2009; Lindenbach *et al.*, 2007; Barzon *et al.*, 2016b; Song *et al.*, 2017).



**Figura 1.4.** Representação simplificada do ciclo replicativo da família *Flaviviridae* (Adaptado de Lindenbach *et al.*, 2013).

#### 1.1.4. Características das proteínas virais

##### 1.1.4.1. Proteínas estruturais

A primeira proteína a ser produzida, C, é altamente básica e apresenta uma massa molecular de  $\approx 11\text{kDa}$ . A proteína nascente possui uma cauda hidrofóbica C-terminal, servindo como péptido sinal na translocação de prM no RE. Após translocação de prM, a cauda é clivada por uma protease viral e uma peptidase sinal tornando C uma proteína madura. Esta possui um domínio interno hidrofóbico, o qual medeia a sua dimerização e a interação com as membranas intracelulares, aquando da formação do invólucro viral (Lindenbach *et al.*, 2013; Barzon *et al.*, 2016b). Apesar de existir pouca informação sobre o reconhecimento e recrutamento do RNA viral pela proteína C, pensa-se que a região N-terminal, juntamente com resíduos carregados da região C-terminal, estejam envolvidos no processo de ligação ao RNA genómico, permitindo, consequentemente, a encapsidação das partículas virais em formação (Samsa, 2009; Lindenbach *et al.*, 2013).

A glicoproteína M ( $\approx 8\text{kDa}$ ) encontra-se inserida no invólucro lipídico e resulta da clivagem de prM ( $\approx 26\text{kDa}$ ), permitindo a maturação dos viriões e originando o segmento pr. A proteína prM colabora no processo em que a glicoproteína E adquire a sua estrutura

tridimensional correta (*folding*) e na heterodimerização com a mesma. Além disso, a presença de prM, num ambiente com pH baixo, previne o rearranjo estrutural de E durante a sua passagem pela via secretória. O segmento pr não se dissocia imediatamente do virião após o processo de clivagem, de forma a proteger o péptido de fusão na proteína E do microambiente celular. Apenas no ambiente extracelular ocorre a completa dissociação, originando-se assim viriões maduros (Smit *et al.*, 2011; Lindenbach *et al.*, 2013).

A glicoproteína E, com  $\approx 53\text{kDa}$ , é a principal proteína de superfície nos viriões de flavivírus. Esta sofre mudanças conformacionais ao longo do ciclo replicativo permitindo a transição do virião entre as suas várias formas (imatura e madura) e modo de fusão ativado. Ela medeia a ligação da partícula viral à superfície da célula hospedeira através da interação com os recetores celulares e a fusão do invólucro viral com a membrana endossomal, permitindo a libertação da nucleocápside no citoplasma (Perera *et al.*, 2008; Lindenbach *et al.*, 2013).

#### **1.1.4.2. Proteínas não-estruturais**

A glicoproteína NS1, com  $\approx 46\text{kDa}$ , encontra-se no interior de células infetadas junto dos complexos replicativos, estando envolvida na fase inicial da replicação viral através da interação com a proteína NS4A. Além disso encontra-se na forma secretada nas células de mamíferos (associada em hexâmeros), e mesmo associada à superfície celular sob a forma de dímeros, desempenhando um papel importante tanto na progressão da infeção como na indução da resposta humoral. A forma secretada de NS1 acumula-se no soro e tecido, constituindo um bom biomarcador de diagnóstico de infeção (Muller e Young, 2013; Lindenbach *et al.*, 2013). A proteína NS1 de ZIKV é semelhante estruturalmente com os outros flavivírus, no entanto, apresenta um potencial electrostático diferente na interface envolvida nas ligações com fatores do hospedeiro, podendo assim afetar a patogénese e imunogenicidade do vírus (Song *et al.*, 2016; Barzon *et al.*, 2016b).

A NS2A, com  $\approx 22\text{kDa}$ , é uma proteína hidrofóbica, componente do complexo de replicação viral, e modula a resposta antiviral do hospedeiro. A NS2A também pode apresentar uma forma truncada, designada NS2A $\alpha$ , que está envolvida na montagem do vírus (Lindenbach *et al.*, 2013; Barzon *et al.*, 2016b).

A NS2B, com  $\approx 14\text{kDa}$ , é uma proteína transmembranar que forma um complexo estável com a proteína NS3. O domínio hidrofílico central atua como cofator na NS2B-

NS3 protease serina. O domínio hidrofóbico terminal atua como âncora para o complexo NS2B-NS3 na membrana do RE (Lindenbach *et al.*, 2013; Barzon *et al.*, 2016b).

A NS3, com  $\approx 70$ kDa e dois distintos domínios globulares conectados por um pequeno conjunto de aminoácidos (*linker*), é uma enzima multifuncional com atividades de protease, helicase (RNAhel) e trifosfatase (RTPase) de RNA e nucleósido trifosfatase (NTPase). O domínio N-terminal é responsável pela atividade proteásica que, em conjunto com NS2B, clivam as junções NS2A-NS2B, NS2B-NS3, NS3-NS4A e NS4B-NS5 na poliproteína viral, bem como locais internos em NS2A e NS3, gerando ainda o C-terminal das proteínas C madura e NS4A. A protease NS2B-NS3 de ZIKV encontra-se como um dímero compactado, não se verificando esta estrutura nas proteases dos outros flavivírus. O domínio C-terminal tem atividade de RNAhel, estando envolvido na separação de cadeias (intermediário da replicação) e estruturas secundárias de RNA formadas durante a replicação viral e na remoção de proteínas ligadas ao RNA viral. Este domínio também é responsável pela hidrólise de nucleósidos trifosfatados (atividade de NTPase) e desfosforilação da extremidade 5' antes da adição da estrutura *cap* (atividade de RTPase) (Lindenbach *et al.*, 2013; Lei *et al.*, 2016; Barzon *et al.*, 2016b). Em ZIKV, a RNAhel parece encontrar-se estruturada como um monómero, ao contrário de DENV, funcionando na forma dimérica (Tian *et al.*, 2016). A proteína NS3 constitui um dos alvos prioritários para o desenvolvimento de novas drogas com atividade antiviral, devido ao seu papel crucial no ciclo replicativo viral e no processamento da poliproteína (Tian *et al.*, 2016).

A NS4A e NS4B são proteínas transmembranares hidrofóbicas, com  $\approx 16$ kDa e  $\approx 27$ kDa, respetivamente, com funções essenciais na replicação viral, onde modulam as alterações na membrana do RE necessárias na formação dos complexos replicativos (Zou *et al.*, 2015; Barzon *et al.*, 2016b). Estas proteínas aparentam interagir com NS3 atuando como cofatores (Barzon *et al.*, 2016b).

Finalmente, a proteína NS5, com  $\approx 100$ kDa, é a maior e mais conservada das proteínas virais. Na região N-terminal apresenta um domínio com atividade de metiltransferase (NS5MTase) o qual está envolvido na formação da estrutura *cap* na extremidade 5' do RNA genómico viral. Por outro lado, na região C-terminal a presença de um domínio com atividade de polimerase de RNA dependente de RNA (RpRd) é responsável pela síntese do RNA viral (Lindenbach *et al.*, 2013; Barzon *et al.*, 2016b).

### 1.1.5. Diversidade genética

O estudo filogenético realizado por Faye *et al.* (2014), com base na sequenciação dos genes que codificam a glicoproteína E e a proteína NS5 de 43 isolados de ZIKV, obtidos entre 1947 e 2007 (na África, Ásia e Oceânia), sugeriu a origem de ZIKV por volta de 1920 no Uganda (África Oriental).

Até ao presente, vários estudos filogenéticos (Haddow *et al.*, 2012; Baronti *et al.*, 2014; Grard *et al.*, 2014; Liu *et al.*, 2016; Ellison *et al.*, 2016) levados a cabo com base na análise da sequência completa/parcial do genoma de diferentes estirpes de ZIKV permitem corroborar a existência de duas linhagens. Uma delas parece ter origem africana e a outra asiática, com a linhagem Africana estruturada em dois grupos. De facto, as árvores filogenéticas revelam um agrupamento (*cluster*) de estirpes do Uganda e da Nigéria, formando assim 3 genótipos: África Ocidental (*cluster* nigeriano), África Oriental (*cluster* ugandês) e Asiático (Faye *et al.*, 2014; Musso e Gubler, 2016; Weaver *et al.*, 2016; Lanciotti *et al.*, 2016). As estirpes epidémicas (ilha de Yap, PF e continente americano), surgidas recentemente, encontram-se na linhagem asiática.

### 1.1.6. Transmissão

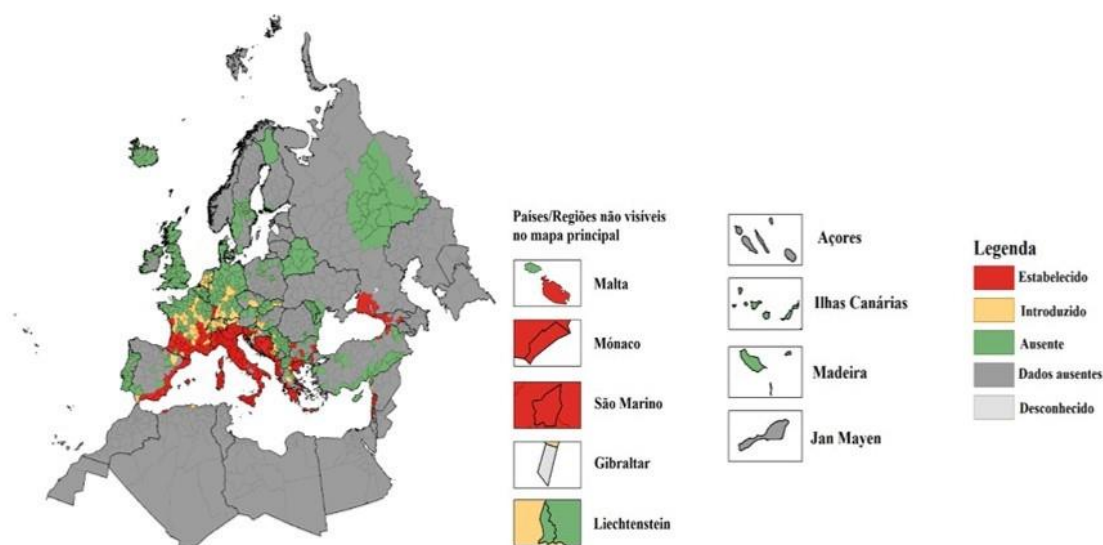
#### 1.1.6.1. Transmissão por mosquitos

ZIKV é principalmente transmitido pela picada de mosquitos. Apresenta dois distintos ciclos de transmissão, um ciclo silvático, onde o vírus é mantido entre primatas não humanos e mosquitos arbóreos nas florestas, e um ciclo urbano, implicado na manutenção do vírus entre humanos e mosquitos urbanos (Weaver *et al.*, 2016; Song *et al.*, 2017). Ocasionalmente, os humanos podem ser acidentalmente infetados por mosquitos arbóreos.

No ciclo urbano, a transmissão é principalmente mediada pelo *Aedes aegypti* e *Aedes albopictus* (Grard *et al.*, 2014; Guerbois *et al.*, 2015), sendo que o primeiro é considerado o vetor primário deste vírus (Song *et al.*, 2017). Estes dois artrópodes encontram-se amplamente distribuídos nas zonas tropicais e subtropicais (ECDC, 2017). No caso de *Aedes albopictus*, a sua presença estende-se a zonas temperadas (Kraemer *et al.*, 2015).

Ocasionalmente, ZIKV foi isolado ou detetado em mosquitos de outros géneros, principalmente, em mosquitos do género *Culex*, pelo que estes têm sido sugeridos na transmissão de ZIKV. No entanto, são necessários estudos adicionais para melhor definir

a competência vetorial para ZIKV e assim determinar outros mosquitos vetores (Baud *et al.*, 2017).



**Figura 1.5.** Distribuição conhecida de *Aedes albopictus* na Europa a abril de 2017 (Adaptado de <https://ecdc.europa.eu/en/publications-data/aedes-albopictus-current-known-distribution-europe-april-2017>).

#### 1.1.6.2. Outros modos de transmissão de ZIKV

Apesar da transmissão de ZIKV aos humanos ocorrer, na maioria dos casos, como consequência da picada de mosquitos, a transmissão sexual, intrauterina e perinatal e por transfusão de sangue foi igualmente já confirmada (Barzon *et al.*, 2016b). Apesar de demonstrados, ainda não se tem a noção da importância epidemiológica destes modos de transmissão.

Vários casos de infecção por ZIKV através de transmissão sexual foram descritos até agora, sendo documentados em cidadãos de países nos quais o ZIKV não circula de forma endêmica, que adquiriram a infecção após relações sexuais com parceiros infetados que retornaram de áreas endêmicas para este vírus. A transmissão é possível a partir de indivíduos que apresentem infecções assintomáticas e sintomáticas através de relação sexual genital e anal e por transmissão do sexo masculino para masculino, masculino para feminino (mais frequentemente) e feminino para masculino (Foy *et al.*, 2011; Deckard *et al.*, 2016; Davidson *et al.*, 2016). A transmissão via outros fluídos corporais não foi demonstrada, no entanto, partículas infecciosas foram encontradas na urina e saliva (Rozé *et al.*, 2016; Barzon *et al.*, 2016a). Transmissão perinatal foi reportada durante o surto na PF em 2013 (Besnard *et al.*, 2014). Transmissão intrauterina foi confirmada durante o

surto no Brasil. O RNA viral foi detetado no líquido amniótico de mulheres grávidas com sintomas de infecção por ZIKV (Calvet *et al.*, 2016). Mais tarde, foi detetado na placenta, cérebro ou soro de fetos/recém-nascidos com microcefalia e de abortos (Mlakar *et al.*, 2016; van der Eijk *et al.*, 2016). ZIKV é possivelmente transmitido por transfusão de sangue de doadores infetados, com a detecção de RNA viral, primeiramente, em doadores assintomáticos no surto da PF (Musso *et al.*, 2014a) e depois em 2016, em Porto Rico, num rastreio de doações de sangue (Kuehnert *et al.*, 2016).

#### **1.1.7. Patogénese e tropismo celular**

Os dados disponíveis sobre a patogénese de ZIKV são escassos, com poucos tipos de células para os quais o tropismo de ZIKV foi provado, e a maioria dos estudos centrados apenas na transmissão transplacentária e danos neuronais (Barzon *et al.*, 2016b).

ZIKV é principalmente transmitido pela picada de mosquitos, pelo que a pele, no sítio de inoculação, representa o primeiro local de replicação viral. Recentemente, Hamel e colaboradores (Hamel *et al.* 2015) demonstraram a permissividade à infecção por ZIKV e sua replicação nos fibroblastos dérmicos primários humanos, queratinócitos epidermais e células dendríticas imaturas. O vírus dispersa-se da pele para os nódulos linfáticos, onde é amplificado, resultando em virémia e disseminação pela corrente sanguínea para tecidos periféricos e órgãos (Barzon *et al.*, 2016b).

Estudos centrados na infecção de células neuronais demonstraram a preferência de ZIKV em infetar células progenitoras neuronais humanas (hNPCs), com indução do mecanismo de apoptose. O vírus também infeta neurónios maduros, no entanto, em menor grau. A capacidade de infecção de células progenitoras sugere assim um potencial mecanismo que afeta o neurodesenvolvimento do cérebro (Tang *et al.*, 2016; Onorati *et al.*, 2016). A infecção por ZIKV de culturas de organóides do cérebro humano (culturas em três dimensões que permitem recapitular os eventos iniciais do desenvolvimento cerebral) levou à redução proliferativa e aumento da morte das hNPCs, com decréscimo do tamanho total dos organóides, lembrando assim os casos de fetos microcefálicos (Garcez *et al.*, 2016; Qian *et al.*, 2016; Dang *et al.*, 2016).

A detecção de RNA viral e partículas infecciosas no líquido amniótico, placenta e cérebro de fetos com microcefalia e abortados permitiu suportar a associação entre a infecção por ZIKV e a microcefalia fetal. Além do mais, estudos conduzidos *in vitro* com

células placentárias humanas também mostraram a distinta suscetibilidade a ZIKV de diferentes células, tais como trofoblastos, células endoteliais do feto e macrófagos placentários (células de Hofbauer), dependendo do tempo de gestação e da variação genética de fatores do hospedeiro na placenta (expressão de fatores de restrição e de ligação ao vírus; Bayer *et al.*, 2016; Quicke *et al.*, 2016; Tabata *et al.*, 2016). O vírus foi igualmente identificado em trofoblastos e células endoteliais do feto na placenta, consistente assim com uma via de infecção transplacentária.

Estudos em modelos animais e em humanos naturalmente infectados demonstraram a capacidade de persistência de partículas virais e RNA viral de ZIKV em vários fluídos corporais (sêmen, saliva, lágrimas e urina), olhos, cérebro e aparelhos reprodutores masculino e feminino (Song *et al.*, 2017). Vários estudos em murganhos demonstraram a capacidade de infecção do aparelho reprodutor masculino (Chan *et al.*, 2016; Ma *et al.*, 2016; Govero *et al.*, 2016). Um estudo *in vitro* descreveu a suscetibilidade dos fibroblastos uterinos humanos a ZIKV (Chen *et al.*, 2016). Este estudo permite, assim, sugerir uma contribuição da infecção uterina/vaginal na transmissão heterossexual.

Estudos em primatas não humanos permitiram recapitular os resultados demonstrados em murganhos. Após inoculação intravenosa, o RNA viral foi recuperado do sangue, saliva, urina, nódulos linfáticos, coração, baço, útero, vagina e tecido gastrointestinal (Coffey *et al.*, 2017). O estudo de infecção em macacos rabo de porco fêmea, que se encontravam grávidas, resultou no desenvolvimento de lesões cerebrais nos fetos (Adams Waldorf *et al.*, 2016).

#### **1.1.8. Apresentação clínica e complicações da infecção por ZIKV**

A infecção por ZIKV apresenta um quadro clínico não específico, pelo que é frequentemente não detetada ou mal diagnosticada como DENV ou infecção a chikungunya (CHIKV; Barzon *et al.*, 2016b; Baud *et al.*, 2017). A percentagem de infecções assintomáticas encontra-se reportada à volta de 80% (Duffy *et al.*, 2009). Os sintomas ocorrem após um período de incubação de 3 a 12 dias e a doença é usualmente ligeira e auto-limitada (Barzon *et al.*, 2016b). Apenas nos surtos recentes foram notificados sintomas e sequelas graves, nomeadamente, a SGB em adultos e a síndrome congénita em fetos e recém-nascidos (Song *et al.*, 2017).



Em indivíduos sintomáticos, os sintomas comuns são exantema maculopapular pruriginoso, febre baixa ( $<38,5^{\circ}\text{C}$ ), artralgia, conjuntivite não purulenta, mialgia e cefaleia (Barzon *et al.*, 2016b; Song *et al.*, 2017). Os sintomas surgem juntamente com a viremia e desaparecem espontaneamente normalmente ao final de uma semana. No entanto, as artrálgias podem persistir durante um mês (Foy *et al.*, 2011). Em áreas de co-circulação de ZIKV, DENV e CHIKV não foi relatado aumento de incidência de doença severa, sugerindo a ausência de efeitos sinérgicos (Dupont-Rouzeyrol *et al.*, 2015; Sardi *et al.*, 2016).

A SGB foi a primeira complicação grave notificada da infecção por ZIKV, em adultos (Baud *et al.*, 2017). É uma doença autoimune, geralmente desencadeada por uma infecção na qual a resposta imune atinge parte do sistema nervoso periférico, causando paralisia flácida aguda (Barzon *et al.*, 2016b; Creange *et al.*, 2016; Song *et al.*, 2017). Atualmente, o mecanismo pelo qual a infecção por ZIKV provoca SGB é desconhecido (Song *et al.*, 2017).

Uma potencial associação entre a infecção por ZIKV e uma síndrome congênita foi identificada através do aumento de casos de microcefalia notificados no Brasil (Kleber de Oliveira *et al.*, 2016; Microcephaly Epidemic Research Group, 2016). Uma análise retrospectiva conduzida na PF permitiu associar temporalmente a infecção de ZIKV e o aparecimento de malformações congênitas afetando o sistema nervoso central. O primeiro trimestre foi sugerido como o período de maior risco para o desenvolvimento de anomalias (Cauchemez *et al.*, 2016). As principais anomalias observadas nos fetos e recém-nascidos incluem microcefalia, ventriculomegalia, calcificações cerebrais e placentárias, atenuação das circunvoluções cerebrais e atrofia cerebral (Vouga e Baud., 2016; Lockwood *et al.*, 2017). Os estudos realizados indicam apenas as complicações observadas no momento, sendo que as sequelas a longo prazo ainda são desconhecidas. (Barzon *et al.*, 2016b; Baud *et al.*, 2017). O papel de co-fatores na síndrome congênita associada a infecção por ZIKV (exemplo: outras infecções, alterações genéticas e uso de álcool) não se encontra identificado, necessitando também de investigação adicional (Rasmussen *et al.*, 2016).

Atualmente, ZIKV surge como um agente patogénico humano altamente relevante, não só pela possibilidade que demonstrou de estar na origem de surtos generalizados, como pela sua associação a síndrome congênita e a SGB, tornando assim evidente, o quão

importante são os estudos centrados neste vírus. Contudo, até ao presente, pouco se conhece sobre a estratégia replicativa utilizada pelo ZIKV nas células permissivas, que tipo de interações são estabelecidas com a maquinaria celular durante o ciclo replicativo, e de que maneira a fisiologia celular e a presença de outros vírus podem influenciar estas interações. Assim, é necessário o desenvolvimento de ferramentas básicas, nomeadamente, de mutantes virais, anticorpos específicos, ou proteínas virais recombinantes a partir das quais possa ser feita a determinação da sua estrutura tridimensional após cristalização. Adicionalmente, estas últimas podem ainda ser utilizadas em ensaios de caracterização da sua atividade enzimática (em particular NS3 e NS5), visando o rastreio *in vitro* de compostos que as possam especificamente inibir, e que possam ser explorados no sentido da produção de fármacos antivirais. Dado que estes objetivos assentam na produção de proteínas virais em quantidade e com elevado grau de pureza e solubilidade, *Escherichia coli* (*E. coli*) surge como um dos possíveis sistemas onde a sua produção pode ser executada.

## **1.2. Expressão de proteínas recombinantes**

A produção de proteínas recombinantes prosperou quando se tornou possível a exploração de sistemas biológicos, frequentemente microbianos, para a produção de proteínas heterólogas, sendo utilizada em diversas áreas, nomeadamente na biofarmacêutica (terapêutica e diagnóstico), agricultura e produção enzimática para a indústria alimentar, têxtil e de detergentes, papel, celulose, polímeros e plásticos. Anterior à utilização dos sistemas microbianos para a produção de proteínas *in vivo*, estas eram derivadas de tecidos animais e vegetais, resultando na sua baixa disponibilidade e preços elevados. Com a utilização de microrganismos de multiplicação rápida e simples e de fácil manipulação genética, os sistemas microbianos evidenciaram-se economicamente favoráveis em relação ao uso de plantas e animais (Demain e Vaishnav, 2009).

A produção de proteínas recombinantes resulta das seguintes etapas: obtenção do DNA codificante de interesse, sua clonagem num vetor de expressão, transformação do hospedeiro/sistema de expressão escolhido, indução e purificação e caracterização da proteína (Rosano e Ceccarelli, 2014). A expressão de proteínas recombinantes é influenciada por vários parâmetros e, por isso, é difícil prever o rendimento de produção,

a solubilidade das proteínas recombinantes e/ou a qualidade conformacional (atividade proteica) (Demain e Vaishnav, 2009).

Existe uma grande variedade de sistemas que podem ser utilizados para a expressão heteróloga de proteínas recombinantes, por isso, a sua escolha é determinada por vários fatores tais como (i) quais os requisitos nutricionais do sistema de expressão, (ii) qual o rendimento de expressão proteica, (iii) qual a localização celular pretendida para o produto a ser expresso (intracelular ou extracelular), (iv) eventual necessidade de modificação pós-traducional, (v) a solubilidade e qualidade conformacional das proteínas produzidas, e (vi) o custo associado (Makrides, 1996). Tendo em conta estes fatores, a escolha reflete-se na utilização de um dos inúmeros sistemas de origem procariota (incluindo bactérias) ou eucariota (incluindo leveduras, fungos filamentosos, plantas, células de insetos e mamíferos) (Demain e Vaishnav, 2009; Rosano e Ceccarelli, 2014). Destacam-se como principais sistemas de origem procariota, as bactérias *E. coli*, *Bacillus megaterium* e *Bacillus subtilis*, e eucariota, as leveduras *Pichia pastoris* e *Saccharomyces cerevisiae*, os fungos filamentosos *Aspergillus niger* e *Aspergillus awamori*, as plantas *Nicotiana benthamiana* (planta do tabaco) e *Arabidopsis thaliana*, as linhas celulares de insetos *Spodoptera frugiperda* (Sf9) e *Trichoplusia ni* (Hi5), às quais é realizada a infeção com baculovírus (vetor de expressão), e as linhas celulares de mamíferos, derivadas de ovário de hamster chinês (CHO) e de rim embrionário humano (HEK293) (Demain e Vaishnav, 2009; He *et al.*, 2014).

### **1.2.1. Sistema de expressão bacteriano: *Escherichia coli***

A bactéria Gram-negativa *E. coli* é, de entre os vários hospedeiros de expressão de origem procarionte disponíveis, a mais utilizada na produção de proteínas recombinantes (Terpe, 2006; Demain e Vaishnav, 2009). Além da sua caracterização genética e bioquímica se encontrar bem compreendida, permitindo a fácil manipulação genética (Demain e Vaishnav, 2009), *E. coli* apresenta outras vantagens, que fazem de si o hospedeiro de eleição para a produção de proteínas recombinantes em hospedeiros heterólogos. Estas características incluem uma fácil sub-cultura e manutenção, rápido crescimento (tempo de duplicação, entre 20-30min), bem como a existência de inúmeras estirpes nas quais a expressão proteica pretendida pode ser otimizada e da qual, normalmente, esta é caracterizada por um elevado rendimento, conseguido a um custo

relativamente baixo (Yin *et al.*, 2007; Demain e Vaishnav, 2009; Rosano e Ceccarelli, 2014).

Por outro lado, *E. coli* apresenta um conjunto de limitações, nomeadamente quando se pretende a produção de proteínas de células eucarióticas ou dos vírus que as infetam. Este sistema não permite a ocorrência, por exemplo, de modificações pós-traducionais (N- e O- glicosilação, acilação dos ácidos gordos, fosforilação e estabelecimento de ligação dissulfeto), por vezes necessárias para a correta funcionalidade proteica (Yin *et al.*, 2007). Para além disso, a produção pode ser limitada devido ao desajuste na utilização de codões por *E. coli*. O desajuste ocorre quando a frequência de utilização de codões sinónimos na sequência nucleotídica de interesse difere da do hospedeiro em que estas se pretendem produzir de forma heteróloga. Em cada célula, a população de RNA de transferência (tRNAs) reflete, aproximadamente, a frequência de utilização da população de codões. Portanto, se o mRNA de interesse contém codões raros em *E. coli*, no momento de síntese das proteínas recombinantes, ocorre a diminuição da quantidade de tRNAs de baixa abundância, levando à incorporação de aminoácidos erróneos e/ou o truncamento do polipéptido desejado, afetando os níveis de expressão e/ou a qualidade conformacional das proteínas recombinantes produzidas (Dong *et al.*, 1996; Gustafsson *et al.*, 2004; Terpe, 2006; Rosano e Ceccarelli, 2014). Por outro lado, a expressão de proteínas recombinantes pode ser limitada pelo facto de estas poderem afetar determinadas funções essenciais no hospedeiro de expressão, e assim interferir na proliferação e homeostasia de *E. coli*, resultando numa diminuição da taxa de crescimento e densidade celular, e/ou na lise celular (Doherty *et al.*, 1993; Dong *et al.*, 1995; Rosano e Ceccarelli, 2014). A expressão de proteínas recombinantes no microambiente de *E. coli* difere do seu local de origem. Por isso mesmo, parâmetros como pH, osmolaridade, potencial redox, cofatores disponíveis e mecanismos de *folding* também são diferentes. Alterações nestes parâmetros podem levar a instabilidade e agregação proteica, resultando na formação de corpos de inclusão (CI; ver secções seguintes) (Hartley e Kane, 1988; Carrió e Villaverde, 2001; Rosano e Ceccarelli, 2014).

### 1.2.1.1. Estratégias de otimização da produção de proteínas recombinantes em *E. coli*

O nível de expressão proteica é afetado pela força do sistema de expressão, a qual é definida pela presença/utilização, no vetor de expressão, de um promotor de transcrição forte, frequentemente reconhecido por uma polimerase de RNA específica e facilmente indutível (são destes exemplos os promotores *T7-lac*, *lac*, *tac*, *trc* e *araP<sub>BAD</sub>*), pelo número de cópias por célula do vetor de expressão (normalmente um plasmídeo) e pelas condições ambientais que afetam o processo de expressão, como a temperatura e a taxa de multiplicação celular (Hellmuth *et al.*, 1994; Keasling, 1999; Rodríguez-Carmona *et al.*, 2012; Marschall *et al.*, 2016).

Tal como acima é referido, a expressão de proteínas recombinantes de origem heteróloga pode ser limitada pela capacidade de provocarem toxicidade sobre o hospedeiro de expressão. O problema pode ser contornado regulando firmemente o nível de expressão basal, ou seja, a expressão na ausência de indutor (Rosano e Ceccarelli, 2014). Se utilizar vetores de expressão com promotores baseados no operão *lac* pode-se adicionar glucose ao meio de cultura, e/ou conjugar este promotor com múltiplas sequências operadoras, às quais se pode ligar o repressor LacI (Studier, 2005; Rosano e Ceccarelli, 2014). Para os vetores de expressão baseados no promotor de transcrição do fago T7 (promotor reconhecido pela polimerase de RNA do fago T7, o qual não é reconhecido pela polimerase de RNA de *E. coli*), o controlo pode ser efetuado, por exemplo, pela utilização de estirpes que co-expressam um gene que codifica a lisozima do fago T7, a qual é inibidora natural da polimerase de RNA do fago T7. A enzima liga-se à polimerase impedindo a transcrição a partir do promotor *T7*, levando à redução do nível de expressão proteica basal, e permitindo a estabilização de plasmídeos que expressam proteínas tóxicas. O gene que codifica a lisozima do fago T7 está disponível num plasmídeo compatível (pLysS ou pLysE) com o sistema de expressão utilizado (Studier e Moffatt, 1986; Moffatt e Studier, 1987; Studier, 1991; Rosano e Ceccarelli, 2014). Por outro lado, pode optar-se pela utilização de vetores de expressão com promotores firmemente regulados, nomeadamente, o promotor *araP<sub>BAD</sub>*, o qual é controlado pela arabinose. No limite, para contornar o problema de toxicidade associado à expressão basal utilizando os vetores de expressão baseados no promotor de transcrição do fago T7, a expressão de proteínas heterólogas pode ser iniciada apenas quando a

polimerase de RNA do fago T7 é transferida através de infecção de *E. coli* pelo fago  $\lambda$ CE6. Assim o gene codificante da proteína tóxica pode ser primeiramente estabilizado em *E. coli* e após infecção é obtida uma elevada quantidade de proteína desejada (Miao *et al.*, 1993). Outra estratégia para contornar o problema de toxicidade passa pela diminuição do número de cópias por célula do vetor de expressão (Rosano e Ceccarelli, 2014). O nível de toxidade de uma proteína recombinante, em determinados casos, torna-se aparente quando um certo limite de tolerância do hospedeiro é alcançado e superado (Rosano e Ceccarelli, 2014). Nesses casos é necessário a manipulação do nível de expressão proteica através de sistemas de expressão nos quais o nível de expressão é ajustável (Wagner *et al.*, 2008), da utilização de estirpes adaptadas à expressão de proteínas tóxicas (Miroux e Walker, 1996) ou do encaminhamento da proteína de interesse para o periplasma ou meio de cultura (Mergulhão *et al.*, 2005). No limite, o gene de interesse pode ser inserido no cromossoma de *E. coli* de forma a limitar a dosagem génica (Baneyx, 1999).

As estratégias de correção do problema de desajuste na utilização de codões passam pela otimização da sequência codificante da proteína de interesse através da sua mutagénesse dirigida ou da síntese *de novo* (artificial), de forma a gerar variantes alélicas com codões reconhecidos pelo conjunto de tRNAs do hospedeiro de expressão, sem modificar a sequência primária da proteína recombinante. Por outro lado, pode ser aumentada a disponibilidade de tRNAs de baixa abundância no sistema, através da utilização de hospedeiros de expressão com genes que codificam tRNAs que reconhecem codões raramente utilizados pelos mesmos. Os genes estão geralmente disponíveis num plasmídeo compatível (Gustafsson *et al.*, 2004; Sørensen e Mortensen, 2005b; Rosano e Ceccarelli, 2014).

Se a quantidade de proteína recombinante produzida permanecer baixa com as estratégias acima propostas, pode optar-se pelo incremento do rendimento volumétrico da proteína de interesse através do aumento da densidade celular, o qual é obtido, pela alteração da composição do meio de cultura (Blommel *et al.*, 2007), através do seu suplemento com extrato de levedura, peptona e catiões divalentes (Studier, 2005) e proporcionando uma melhor aeração através de agitação vigorosa (Losen *et al.*, 2004; Rosano e Ceccarelli, 2014).

As estratégias aplicadas para evitar a formação de CI são variadas. A baixa solubilidade proteica pode ser melhorada utilizando um parceiro de fusão, ou seja, uma proteína altamente solúvel e estável que promove a solubilidade de proteínas alvo, de forma a serem reunidas as condições necessárias para a aquisição de uma estrutura tridimensional correta (Hammarstrom *et al.*, 2002). Os parceiros de fusão disponíveis são inúmeros, mas frequentemente são utilizados a proteína de ligação a maltose (MBP), a proteína de terminação/anti-terminação NusA, a glutathione S-transferase (GST), a tiorredoxina (Trx), a pequena proteína modificadora semelhante a ubiquitina (SUMO) e, mais recente, a proteína de ligação ao cálcio de 8kDa (Fh8). O estabelecimento de ligações dissulfeto é necessário para a conformação tridimensional biológica ativa de muitas proteínas recombinantes. O estabelecimento errôneo destas ligações pode levar a agregação e formação de CI (Rosano e Ceccarelli, 2014). Em *E. coli*, a formação destas ligações ocorre no periplasma (Messens e Collet, 2006), pelo que uma das estratégias passa pelo encaminhamento da proteína de interesse para o periplasma (Mergulhão *et al.*, 2005). Por outro lado, a expressão no citoplasma pode ser alcançada utilizando estirpes modificadas geneticamente na via da tiorredoxina e glutathione/glutarredoxina, de modo a tornar o citoplasma um ambiente menos redutor, e assim favorecendo o estabelecimento de ligações dissulfeto (Derman *et al.*, 1993; Stewart *et al.*, 1998). Se necessárias modificações pós-traducionais além do estabelecimento de ligação dissulfeto é necessário alterar o sistema de expressão. A aquisição de uma estrutura tridimensional específica nas proteínas depende da sua correta estruturação tridimensional, a qual, depende de mecanismos de controlo de qualidade proteica (Rosano e Ceccarelli, 2014). O elevado nível de expressão de proteínas recombinantes pode resultar na saturação destes mecanismos, nomeadamente, das chaperones moleculares, levando a agregação e formação de CI. As estratégias de correção passam pela co-expressão de chaperones como GroEL-GroES, DnaK/DnaJ/GrpE e *trigger factor* (TF), as quais são expressas a partir de plasmídeos, individualmente ou em combinações (Nishihara *et al.*, 1998; Nishihara *et al.*, 2000; de Marco *et al.*, 2007). Por outro lado, o meio de cultura pode ser suplementado com osmólitos (designados de chaperones químicos), como prolina, glicina betaína e trealose (de Marco *et al.*, 2005). O processo de *folding* proteico pode requerer cofactores específicos, particularmente, iões metálicos (magnésio e centros ferro-enxofre), os quais poderão ser adicionados ao meio de cultura (Sørensen e Mortensen, 2005b). Uma outra

forma para permitir o correto processo de *folding* proteico é através da redução da taxa de síntese proteica, pela diminuição da temperatura de incubação (15-25°C) (Vera *et al.*, 2007) ou utilizando estirpes de expressão com chaperones adaptadas a baixas temperaturas (Ferrer *et al.*, 2004), ou sistemas nos quais o nível de expressão é ajustável (Wagner *et al.*, 2008).

Os CI podem conter a proteína pura e intacta, por isso, pode-se optar pelo *refolding in vitro* da mesma, e assim poderá ser possível a obtenção de proteína na conformação nativa (Rosano e Ceccarelli, 2014).

### 1.3. Objetivos

Tal como referido acima, pouco se conhece sobre a estratégia replicativa que permite ao ZIKV replicar em células permissivas, que tipo de interações são estabelecidas com a maquinaria celular durante o ciclo replicativo, ou de que maneira a fisiologia celular e/ou a presença de outros vírus podem influenciar estas interações. Por isso mesmo é necessário o desenvolvimento de ferramentas básicas que permitam abordar estas questões. Assim foi estabelecido como objetivo principal a otimização da produção/purificação de antígenos virais de ZIKV em *E. coli*, nomeadamente, a proteína estrutural da cápside, C, e um constituinte do complexo replicativo, NS3, em particular, a região C-terminal com atividade enzimática de RNA<sub>h</sub>el (NS3<sub>hel</sub>). Para dar resposta ao objetivo principal apresentado foram estabelecidos os seguintes objetivos específicos:

- Otimizar as sequências nucleotídicas que codificam C e NS3<sub>hel</sub>, de modo a ajustar a preferência de utilização de códons em *E. coli*.
- Amplificar, por reação em cadeia da polimerase antecederida de transcrição reversa (RT-PCR), as regiões da C e NS3<sub>hel</sub> de sequências virais de genótipo selvagem, a partir do genoma viral e utilizando *primers* específicos, com locais de restrição enzimática, para permitir o posterior processo de clonagem.
- Amplificar, por reação em cadeia da polimerase (PCR), as regiões da C e NS3<sub>hel</sub> de sequências nucleotídicas sintéticas otimizadas do genoma de ZIKV, utilizando *primers* específicos, com locais de restrição enzimática, para permitir o posterior processo de clonagem.
- Clonar as sequências virais de genótipo selvagem e versões sintéticas com sequências otimizadas num vetor de expressão, de modo, a possibilitar a expressão



em *E. coli*, como proteínas de fusão com caudas de hexa-histidina (His<sub>6</sub>) na extremidade C-terminal, e assim permitir a sua subsequente purificação por cromatografia de afinidade.

- Determinar as condições ótimas de produção através da análise, em paralelo, da expressão proteica de sequências virais de genótipo selvagem e versões sintéticas com sequências otimizadas em diferentes estirpes de *E. coli*, *BL21(DE3)pLysS* e *Rosetta(DE3)pLysS*, tendo em conta parâmetros, como a temperatura de crescimento, a concentração do indutor, a heterogeneidade da população bacteriana e a cinética de produção.
- Determinar as condições ótimas de solubilidade proteica, tendo em conta parâmetros, como a temperatura de crescimento, o tempo de indução, a fase de crescimento na qual a indução da expressão proteica é efetuada, o uso do aditivo D-sorbitol no meio de crescimento e o mecanismo de lise celular utilizado.
- Purificar C e NS3hel por cromatografia de afinidade em colunas com íons níquel.

## 2. MATERIAIS E MÉTODOS

### 2.1. Otimização das sequências nucleotídicas codificantes de C e NS3hel para ajustamento da preferência de utilização de codões em *E. coli*

As regiões codificantes dos polipéptidos C (proteína da cápside viral) e NS3hel (domínio ao qual é atribuída atividade helicase à proteína NS3) na sequência correspondente ao genoma completo de ZIKV, disponível na base de acesso público (acedido através do portal do *National Center for Biotechnology Information*-NCBI, em <https://www.ncbi.nlm.nih.gov/nucleotide>), com o número LC002520, foram modificadas artificialmente (NZYtech, Portugal) de forma a gerar novas sequências codificantes com codões reconhecidos pelo conjunto de tRNAs de *E. coli*, sem modificar a sequência primária das proteínas C e NS3hel. A otimização das sequências nucleotídicas para ajustamento da preferência de utilização de codões em *E. coli* foi realizada utilizando um algoritmo desenvolvido pela empresa à qual o serviço de síntese foi requisitado, designado, ATGenium. Em LC002520, a região C apresenta as coordenadas 107-472, enquanto que a região NS3hel se estende entre as coordenadas 5681-6031.

O processo de síntese artificial das sequências codificantes pretendidas faz uso da utilização de oligonucleotídeos parcialmente sobreponíveis que, em conjunto, cobrem toda a região a sintetizar, na amplificação do produto final por PCR e na correção de erros que possam ocorrer durante a síntese devido à presença da endonuclease do fago T7. As sequências nucleotídicas sintéticas otimizadas que codificam C e NS3hel foram clonadas nos vetores pHTP0 e pNZY29 (NZYtech, Portugal), respetivamente. Ambos os vetores contêm um gene que confere resistência a ampicilina (Amp). As sequências nucleotídicas sintéticas otimizadas que codificam C e NS3hel encontram-se descritas no Anexo I.

### 2.2. Extração de RNA total

A extração de RNA total, a partir do sobrenadante de cultura de células Vero E6 infetadas por ZIKV (estirpe MR 766), foi efetuada utilizando o sistema comercial *ZR Viral RNA Kit*<sup>TM</sup> (Zymo Research, E.U.A) e seguindo as instruções recomendadas pelo fabricante. O procedimento de extração consistiu num passo de descongelamento de alíquotas de sobrenadante de cultura enriquecido em partículas virais (previamente conservadas a -80°C) à temperatura ambiente, seguido de um passo de lise das partículas

virais e aplicação numa coluna de extração para permitir a adsorção do RNA à coluna. O próximo passo resultou na lavagem da coluna para a remoção de contaminantes. Por fim, o RNA foi eluído por adição de 15µl de água livre de RNases à coluna. O eluato foi dividido em duas alíquotas, de 6µl e 9µl, e estas foram guardadas a -80°C.

### 2.3. Síntese, por transcrição reversa, de DNA complementar, a partir de extratos de RNA total

A síntese de DNA complementar (cDNA) a partir do extrato de RNA total (obtido em 2.2) foi efetuada através da utilização do sistema comercial *NZY First-Strand cDNA Synthesis Kit, separate oligos* (NZYtech, Portugal) e seguindo as instruções recomendadas pelo fabricante. A 6µl do extrato de RNA total (ou a 6µl de água livre de nucleases, utilizado como controlo negativo) foram adicionados 1µl de oligonucleótidos (hexâmeros) sintéticos iniciadores com sequências aleatórias e 1 µl de solução tampão de hibridação (0,5X final), perfazendo um volume total de 8µl. Após 2 a 3 aspirações/expulsões com micropipeta, as misturas foram incubadas a 65°C durante 5min. Os tubos foram mantidos em banho de gelo por igual período de tempo, e de seguida foi preparada uma mistura de reação, a qual incluía 10µl de *NZYRT Master Mix, no oligos 2X* e 2µl de *NZRT Enzyme Mix*, perfazendo um volume total de 12µl. Após homogeneização e centrifugação rápida, a mistura de reação foi adicionada aos 8µl da mistura de RNA/hexâmeros/solução tampão de hibridação, perfazendo um volume total de 20µl. A mistura resultante foi homogeneizada suavemente e foi efetuada uma centrifugação rápida. A reação de transcrição reversa (RT) foi realizada no termociclador *Hybaid PCR Express* (Thermo Fisher Scientific, E.U.A.) de acordo com o programa descrito na Tabela 2.1. Após terminação deste processo, os tubos foram colocados em banho de gelo durante 5min, foi adicionado 1µl de *NZY RNase H* e, por fim, incubados a 37°C durante 20min para degradação do RNA molde. O cDNA foi conservado a -20°C até posterior utilização.

**Tabela 2.1.** Condições utilizadas na reação de RT.

Etapa	Temperatura (°C)	Tempo de duração (min)	Número de ciclos
Hibridação	25	10	1
Extensão	50	30	1
Inativação da enzima	85	5	1

#### 2.4. Delineamento de oligonucleótidos sintéticos iniciadores de reação em cadeia da polimerase

O desenho da sequência de oligonucleótidos sintéticos iniciadores (*primers*) para a amplificação, por PCR, de uma sequência genômica para posterior detecção de ZIKV, a partir do cDNA preparado tal como descrito em 2.3., foi elaborado manualmente, com recurso a um alinhamento múltiplo de sequências nucleotídicas de referência disponíveis no GenBank (acedido através do portal do NCBI, em <https://www.ncbi.nlm.nih.gov/>). A escolha da sequência dos *primers* em questão baseou-se na pesquisa de regiões conservadas entre as várias sequências de referência alinhadas, com recurso ao programa bioinformático *MAFFT* version 7 (disponível em <http://mafft.cbrc.jp/alignment/server/index.html>). Igualmente, o desenho de *primers* para a amplificação das sequências nucleotídicas de genótipo selvagem de ZIKV que codificam C e NS3hel foi elaborado manualmente, com base na sequência de referência LC002520. Por seu lado, o desenho de *primers* para a amplificação das sequências nucleotídicas sintéticas otimizadas que codificam C e NS3hel, a partir dos plasmídeos em que foram inicialmente fornecidas foi também elaborado manualmente, com base nas sequências nucleotídicas descritas no Anexo I. Numa primeira abordagem, o par de *primers* complementar à sequência codificante de C foi desenhado de forma a permitir a amplificação da totalidade da mesma. No entanto, mais tarde o *primer* reverso (*primer R*) foi redesenhado de forma a eliminar uma região no extremo 3' da região que codifica uma secção transmembranar (hidrofóbica), localizada no extremo carboxilo (C-terminal) da proteína C.

A cada um dos *primers*, na respetiva extremidade 5', foram adicionados nucleótidos não complementares às sequências nucleotídicas a amplificar, e que correspondem a locais de clivagem por enzimas de restrição do tipo II. As sequências de restrição introduzidas eram reconhecidas pelas enzimas *NotI* e *NdeI*. A sequência de restrição da enzima *NdeI* foi adicionada aos *primers* diretos (*primers F*) e a de *NotI* aos *primers* reversos (*primers R*). Para além da sequência de restrição foram também adicionados nucleótidos não complementares, com o intuito de melhorar a eficácia de corte das enzimas de restrição. A estes últimos nucleótidos designou-se como sequência espaçadora.

Relativamente à detecção do genoma de ZIKV e amplificação das regiões da C e NS3hel de genótipo selvagem, toda a informação acerca dos *primers* encontra-se descrita na Tabela 2.2. Toda a informação na Tabela 2.3 é relativa aos *primers* para a amplificação das sequências nucleotídicas sintéticas otimizadas que codificam C e NS3hel, e ao *primer R* redesenhado para a amplificação da região C. Os *primers* foram adquiridos comercialmente à empresa STAB-VIDA (Portugal).

**Tabela 2.2.** Informação relativa ao conjunto de *primers* utilizados nas amplificações, por PCR, de uma sequência genómica (coordenadas 165-661 na sequência de genoma completo, LC002520) para detecção de ZIKV, e das regiões da C e NS3hel de genótipo selvagem.

Região do genoma	Primer (designação)	Sequência <sup>1</sup> (5' → 3')	Tamanho (nt)	Tm <sup>1</sup> (°C)	Posição (LC002520)
—	ZIKV_FW	GAGTAGCCCGTGTAACCCC	20	57,6	165 - 184
—	ZIKV_REV	ATCTGGTTCCACTCCCTCATC	21	56,2	661 - 641
C	Cpwtfwr	CCGGTTAATCATATGAA AAACCCAAAGAAGAAAT CCGG	38	56.3	107 - 132
	Cpwtrev	ATGAGCTGGCGGCCGC TGCCATGGCTGTAGTCAGCAG	37	59.8	472 - 452
NS3hel	Helwtfwr	CCGGTTAATCATATGG GGAAAACAGTTTGGTTC GTTCC	38	56.9	5681 - 5703
	Helwtrev	ATGAGCTGGCGGCCGC ATACATGTACTCATCTCC AGGTT	39	53	6031 - 6009

**Tabela 2.3.** Informação relativa ao conjunto de *primers* utilizados nas amplificações, por PCR, das sequências nucleotídicas sintéticas otimizadas que codificam C e NS3hel, e ao *primer R* redesenhado para a amplificação da região C.

Região do genoma	Primer (designação)	Sequência <sup>1</sup> (5' → 3')	Tamanho (nt)	Tm <sup>1</sup> (°C)
C	Cpmutfwr	CCGGTTAATCATATGAAA AACCCGAAAAAAGAGT GGC	39	57.2
	Cpmutrev	ATGAGCTGGCGGCCGC GCCATTGCAGTCGTCAGCAGC	38	63.6
	Cmutdel2	ATGAGCTGGCGGCCGCA TCCGCGCCGCGCGTTTAC	36	66.1
NS3hel	Helmutfwr	CCGGTTAATCATATGGGT AAAACGGTGTGGTTTGTGC C	38	59.2
	Helmutrev	ATGAGCTGGCGGCCGCA TACATATATTCGTCACCCG GC	38	54.4

<sup>1</sup>Nota: O início da sequência destacada com realce e a negrito corresponde à sequência espaçadora. A sequência destacada com realce, a negrito e sublinhado corresponde à sequência de restrição de reconhecimento da enzima de restrição. O valor da temperatura de dissociação ( $T_m$ ; do inglês *melting temperature*) dos *primers* refere-se à sequência homóloga a ZIKV, *i.e.*, sem a sequência de restrição e o espaçador (a partir da sequência sem marcação), e foi calculada com recurso à ferramenta *OligoAnalyzer 3.1* disponível em <https://eu.idtdna.com/calc/analyser>. NT corresponde a nucleótidos.

## 2.5. Amplificação de sequências virais de ZIKV por PCR

### 2.5.1. Amplificação de uma sequência viral para confirmação da presença de genoma de ZIKV

A confirmação da presença de genomas de ZIKV num extrato de RNA total foi efetuada através da amplificação, por RT-PCR, de uma sequência genómica (coordenadas 165-661 na sequência de genoma completo, LC002520), utilizando o par de *primers* ZIKV\_FW e ZIKV\_REV (Tabela 2.2). Para a amplificação foi preparada uma mistura de reação, a qual incluiu 12,5µl de *NZYTaq 2× Green Master Mix* (1X final; NZYtech, Portugal), 1µl de *primer* ZIKV\_FW (0.4pmol/µl final), 1µl de *primer* ZIKV\_REV (0.4pmol/µl final) e 10µl de água livre de nucleases. À mistura de reação foram adicionados 0,5µl de cDNA (produto resultante da reação de RT descrita em 2.3.) ou 0,5µl de água livre de nucleases (correspondente ao controlo negativo), perfazendo, assim, um volume final de 25µl. A reação de PCR foi realizada no termociclador *Hybaid PCR Express* (Thermo Fisher Scientific, E.U.A.) e as suas condições encontram-se descritas na Tabela 2.4.

**Tabela 2.4.** Condições de reação utilizadas na amplificação, por PCR, para confirmação da presença de genoma de ZIKV.

Etapa	Temperatura (°C)	Tempo de duração	Número de ciclos
<b>Desnaturação inicial</b>	94	2 min	1
<b>Desnaturação</b>	94	30 seg	40
<b>Hibridação</b>	50	30 seg	
<b>Extensão</b>	72	45 seg	
<b>Extensão final</b>	72	10 min	1
<b>Manutenção</b>	4	∞	1

### 2.5.2. Amplificação das regiões da C e NS3hel de sequências nucleotídicas de genótipo selvagem de ZIKV

Para a amplificação, por PCR, das regiões da C e NS3hel de sequências nucleotídicas de genótipo selvagem de ZIKV foram utilizados os pares de *primers* Cpwtfwrdd e Cpwttrev, e Helwtfwrdd e Helwtrev (Tabela 2.2), respetivamente.

A amplificação foi efetuada através da utilização do sistema comercial *Supreme NZYTaQ DNA polymerase* (NZYtech, Portugal). Para a amplificação foram preparadas misturas de reação, as quais incluíam 5µl de solução tampão de reação (1X final), 2,5µl de cloreto de magnésio (MgCl<sub>2</sub>; 2,5mM final), 1,5µl de desoxirribonucleótidos 5'-trifosfatados (dNTPs; cada um a 0,3mM final), 0,4pmol/µl final de cada um dos *primers*, 0,5µl de *Supreme NZYTaQ DNA polymerase* (5U/µl) e 35,5µl de água livre de nucleases. Às misturas de reação foi adicionado 1µl de cDNA (produto resultante da reação de RT descrita em 2.3) ou 1µl de água livre de nucleases (correspondente ao controlo negativo), perfazendo assim um volume final de 50µl. A reação de PCR foi realizada no termociclador *Hybaid PCR Express* (Thermo Fisher Scientific, E.U.A.) ou *T100™* (Bio-Rad, E.U.A.) e as suas condições encontram-se descritas na Tabela 2.5.

### 2.5.3. Amplificação das regiões codificantes da proteína C e do domínio NS3hel de ZIKV, a partir de sequências nucleotídicas sintéticas

Para a amplificação, por PCR, das regiões que codificam a proteína da C e o domínio NS3hel a partir de sequências nucleotídicas de ZIKV sintéticas, otimizadas para expressão génica em *E. coli*, foram utilizados os pares de *primers* Cpmutfwrdd e Cpmutrev, e Helmutfwrd e Helmutrev (Tabela 2.3), respetivamente. Após o ajuste na sequência do *primer R*, para a eliminação da região C-terminal (hidrofóbica) desta proteína, foi utilizado o par de *primers* Cpmutfwrdd e Cmutdel2 (Tabela 2.3). A amplificação foi efetuada através da utilização do sistema comercial *Supreme NZYTaQ DNA polymerase*. O procedimento utilizado para a amplificação pretendida correspondeu ao descrito em 2.5.2. No entanto, neste caso em concreto, a reação de amplificação foi iniciada adicionando às misturas de reação, 1µl de uma diluição de 1/10 de uma solução contendo o plasmídeo pHTP0 ou pNZY29 onde foi clonada a sequência nucleotídica otimizada da região C ou NS3hel, respetivamente. As condições de amplificação encontram-se descritas na Tabela 2.5.

**Tabela 2.5.** Condições de reação utilizadas na amplificação, por PCR, das regiões codificantes de C e NS3hel, a partir de sequências nucleotídicas de genótipo selvagem e versões sintéticas com códons otimizados.

Etapa	Temperatura (°C)	Tempo de duração	Número de ciclos
<b>Desnaturação inicial</b>	95	2 min	1
<b>Desnaturação</b>	95	30 seg	35
<b>Hibridação</b>	65	30 seg	
<b>Extensão</b>	72	45 seg	
<b>Extensão final</b>	72	5 min	1
<b>Manutenção</b>	4	$\infty$	1

## 2.6. Análise por eletroforese em gel de agarose

Os fragmentos de DNA resultantes das reações de amplificação executadas foram detetados através da sua resolução por eletroforese em gel de agarose a 1,5% em solução tampão TAE 0,5X (Anexo II), na presença de 0,5µg/ml de brometo de etídio. Para a análise foi utilizado 1µl de amostra de cada um dos produtos que resultou das reações de PCR. O tamanho do fragmento obtido em cada um dos casos foi estimado a partir da migração relativa de um marcador de fragmentos de massas moleculares conhecidas (*NZYDNA Ladder VI* - NZYtech, Portugal). Para a migração eletroforética foi aplicada uma voltagem de 85V durante  $\approx$  1 hora. A visualização e registo em suporte digital de imagem, sob luz ultravioleta (UV), do produto amplificado foram obtidos através da utilização do sistema *Gel Doc XR System* (Bio-Rad, E.U.A.).

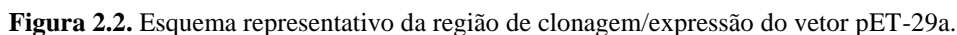
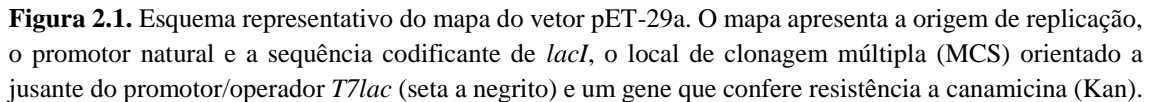
## 2.7. Purificação dos produtos de PCR

A purificação dos produtos obtidos por PCR (descrito em 2.5.2 e 2.5.3) foi efetuada utilizando o sistema comercial *DNA Clean & Concentrator-5<sup>TM</sup>* (Zymo Research, E.U.A.) e seguindo as instruções recomendadas pelo fabricante. O procedimento de purificação consistiu num passo inicial de adsorção do DNA a uma matriz, seguida de adição de solução tampão de lavagem à coluna para remoção de contaminantes. Por fim, o DNA foi eluído por adição de 15µl de solução tampão de eluição de DNA à coluna.

O rendimento do processo de purificação foi avaliado através da análise de 1µl do extrato obtido, por eletroforese em gel de agarose a 1,5%, tal como descrito em 2.6.



Os fragmentos de DNA amplificados por PCR e purificados (obtidos em 2.7), correspondendo às regiões que codificam C e NS3hel de ZIKV quer usando sequências nucleotídicas sintéticas otimizadas (descrito em 2.5.3.), quer de genótipo selvagem (descrito em 2.5.2) foram clonados no vetor plasmídico pET-29a, usando *E. coli NovaBlue* (Novagen, Alemanha) como hospedeira de clonagem. O mapa e as características do vetor pET-29a encontram-se esquematizados nas Figuras 2.1 e 2.2.



### **2.8.1. Digestão enzimática do vetor de clonagem/expressão pET-29a e dos produtos de PCR**

A construção dos vetores de clonagem/expressão contendo os insertos das regiões codificantes de C e NS3hel de ZIKV, nas suas versões sintética com códons otimizados e de genótipo selvagem, foi iniciada através da realização de clivagem enzimática com endonucleases de restrição. O vetor pET-29a bem como os fragmentos de DNA amplificados por PCR (descrito em 2.5.2 e 2.5.3) e purificados (descrito em 2.7) foram digeridos com as enzimas de restrição *NotI* e *NdeI*. Para a digestão enzimática foram preparadas misturas de reação para os produtos de PCR e vetor num volume final de 30µl ou 50µl, respectivamente, dependendo da quantidade de produto obtido no processo de purificação. Estas incluíam, 14µl de produto de PCR purificado ou 5µl de vetor pET-29a, 3 ou 5µl de solução tampão *Orange* (concentrada 10X; Thermo Fisher Scientific, E.U.A.), 1 ou 2µl de cada enzima de restrição (10U/µl; Thermo Fisher Scientific, E.U.A.) e água livre de nucleases, até perfazer o volume total desejado. As reações de digestão enzimática foram incubadas a 37°C, em banho-maria, durante a noite.

O produto resultante da digestão do vetor pET-29a foi analisado por eletroforese em gel de agarose a 1% usando 1µl da mistura de digestão enzimática, tal como descrito em 2.6. O tamanho do fragmento (plasmídeo linearizado) obtido foi estimado a partir da migração relativa de um marcador de fragmentos de massas moleculares conhecidas (*Lambda – pUC Mix Marker*, 4 – Thermo Fisher Scientific, E.U.A.).

#### **2.8.1.1. Purificação dos produtos resultantes da digestão enzimática**

Anterior à purificação foi efetuada a migração eletroforética do produto resultante da digestão enzimática correspondente ao vetor pET-29a digerido (49µl). Para a mesma foi aplicada uma voltagem de 75V durante  $\approx$  1h a 1h e 30min. A banda correspondente ao fragmento do vetor pET-29a digerido foi excisada do gel com o auxílio de um bisturi sob irradiação UV. De seguida, a purificação do vetor pET-29a digerido foi efetuada utilizando o sistema comercial *Zymoclean™ Gel DNA Recovery Kit* (Zymo Research, E.U.A.), seguindo as instruções recomendadas pelo fabricante. O procedimento de purificação consistiu na pesagem do cubo de gel de agarose removido e adição de um volume de solução tampão de dissolução de agarose (ADB) na proporção 1:3 de gel:ADB. A mistura resultante foi incubada a 55°C durante 10 minutos até uma completa dissolução

da agarose, e aplicada numa coluna de purificação. De seguida foi efetuada a adsorção do DNA plasmídico à coluna e a lavagem da mesma para remoção de contaminantes. Por fim, o DNA plasmídico foi eluído por adição de 15µl de água livre de DNase/RNase à coluna.

A purificação dos produtos de PCR digeridos foi efetuada utilizando o sistema comercial *DNA Clean & Concentrator-5<sup>TM</sup>*. O protocolo foi realizado tal como descrito em 2.7.

O rendimento do processo de purificação foi avaliado pela estimativa da concentração de ácidos nucleicos, por medição da densidade ótica, medida a 260nm (D.O.<sub>260nm</sub>) no equipamento *NanoDrop® ND-1000* (Thermo Fisher Scientific, E.U.A.).

### **2.8.2. Ligação entre o vetor de clonagem/expressão pET-29a e produtos de PCR**

Para a ligação entre o vetor pET-29a e os fragmentos de DNA usados como insertos (razão molar inserto/vetor de 3:1) foram preparadas misturas de ligação, as quais incluíram 1µl de solução tampão de ligação concentrada 10X (Thermo Fisher Scientific, E.U.A.), 1,5µl de vetor pET-29a digerido e purificado (73ng/µl) (descrito em 2.8.1 e 2.8.1.1, respetivamente), 4µl de produto de PCR digerido e purificado (descrito em 2.8.1 e 2.8.1.1, respetivamente), 1µl de DNA ligase do fago T4 (5U/µl) (Thermo Fisher Scientific, E.U.A.) e água livre de nucleases para perfazer um volume final de 10µl. A mistura de ligação foi incubada durante 3h à temperatura ambiente. Esta foi posteriormente utilizada na transformação de células competentes de *E. coli NovaBlue*.

### **2.8.3. Preparação de células competentes de *E. coli***

O procedimento foi efetuado com base no protocolo ligeiramente modificado, descrito inicialmente por Chung e colaboradores (Chung *et al.*, 1989). A preparação de células competentes foi iniciada através da inoculação de uma suspensão celular da estirpe *NovaBlue* (pura e concentrada) conservada a -80°C em glicerol, em 3ml de meio Luria-Bertani líquido (LB; Sambrook *et al.*, 1989; Anexo II; NZYtech, Portugal) seguindo uma incubação a 37°C com agitação (220rpm) durante a noite. Após esse período, 150µl de cultura saturada obtida foram diluídos em 15ml de meio LB líquido (diluição 1:100), e novamente incubados nas condições de crescimento celular descritas

acima até ter sido atingida uma densidade ótica, medida a 600nm (D.O.<sub>600nm</sub>), a variar entre 0,3-0,5. As células bacterianas foram colhidas por centrifugação a 3000g, durante 15min, a 4°C. Após decantação do sobrenadante, o sedimento celular foi ressuspensionado em 500µl de solução de transformação e armazenamento (TSS) fria (Anexo II).

#### **2.8.4. Transformação de *E. coli* por choque térmico**

A tubos de 1,5ml frios contendo 100µl de células ressuspensas em TSS (descrito em 2.8.3) foram adicionados 5µl de cada mistura de ligação (descrito em 2.8.2). O mesmo procedimento foi efetuado para o controlo negativo, tendo-se adicionado 5µl de água destilada estéril. As suspensões celulares obtidas foram incubadas no gelo durante 30min, seguida de submissão a um choque térmico a 42°C durante 90seg, e novamente transferidas para gelo durante 5min. Terminada a incubação em gelo foram adicionados 900µl de meio LB líquido aos tubos, seguido de incubação de 1h a 37°C, com agitação suave (70-75rpm). As suspensões bacterianas foram semeadas em placas de Petri com meio LB sólido (Anexo II) suplementado com Kan (30µg/ml). Cada suspensão foi semeada em triplicado, sendo que duas placas foram inoculadas com 100µl cada e a última, com o volume final (aproximadamente 100µl) da ressuspensão do sedimento celular resultante de uma centrifugação a 16000g durante 1min. Por fim foi semeado 100µl de suspensão bacteriana correspondente ao controlo negativo. As placas foram incubadas a 37°C numa estufa, durante a noite (entre 16h a 18h) de forma a permitir o crescimento de colónias bacterianas.

#### **2.8.5. Extração de DNA plasmídico pelo método de lise alcalina**

A extração de DNA plasmídico foi executada através da lise alcalina de suspensões bacterianas, com base no protocolo descrito por Birnboim e Doly (Birnboim e Doly, 1979). Colónias bacterianas isoladas foram repicadas, com um palito estéril, para tubos com 3ml de meio LB líquido suplementado com Kan (30µg/ml), e incubadas durante a noite a 37°C, com agitação (220rpm). Posteriormente, centrifugou-se 1,8ml das culturas bacterianas saturadas obtidas a 16000g durante 1min. Após decantação do sobrenadante, o sedimento celular foi ressuspensionado, com auxílio de um agitador de tipo vórtex, em 300µl de solução TEG (25mM Tris-HCl pH 7,5; 10mM EDTA; 1% (m/v) glucose) fria. De seguida foram adicionados 300µl de solução de lise (150mM NaOH; 1,5% (m/v)

SDS), promovendo a lise bacteriana por inversão suave dos microtubos. Após colocação dos microtubos em gelo durante 5min foram adicionados 300µl de solução de acetato de potássio 3M (pH 5,4), e pela sua inversão suave foi permitida a neutralização do pH dos lisados. Depois de nova colocação dos microtubos em gelo durante 5min procedeu-se à centrifugação dos mesmos a 16000g durante 15min. O sobrenadante foi transferido para novos microtubos para permitir a precipitação dos ácidos nucleicos através da adição de 900µl de isopropanol, e posterior homogeneização por inversão suave. Os ácidos nucleicos precipitados foram colhidos por centrifugação a 16000g durante 30min. O sedimento de ácidos nucleicos foi lavado com 300µl de solução de etanol a 70%, seguido de centrifugação a 16000g durante 5min. Após decantação do sobrenadante, o sedimento foi seco sob vácuo (*Concentrator 5301* – Eppendorf AG, Alemanha) durante 5-10min. Para finalizar, o sedimento seco foi ressuscitado em 50µl de solução tampão TE (10mM Tris-HCl pH 8; 1mM EDTA) com 50µg/ml de RNaseA e incubado a 37°C, em banho-maria, durante 30min, para permitir a digestão do RNA presente.

#### **2.8.6. Análise de moléculas recombinantes de DNA plasmídico**

A confirmação da natureza recombinante das moléculas de DNA plasmídico foi efetuada através da amplificação, por PCR, dos insertos de DNA nelas contidos. Para a amplificação foi utilizada uma mistura de reação pronta a usar, *NZYTaq 2× Green Master Mix*, como descrito em 2.5.1. Como *primers* utilizou-se o par usado nas amplificações do inserto de DNA correspondente (descrito em 2.5.2 e 2.5.3). As misturas de reações incluíam, 12,5µl de *NZYTaq 2× Green Master Mix* (1X final), 0,4pmol/µl final de cada um dos *primers* utilizados na obtenção de cada um dos amplicões a clonar, e 9,5µl de água livre de nucleases. Às misturas de reações foram adicionados 1µl de solução contendo o DNA plasmídico extraído (obtido em 2.8.5), ou 1µl de água livre de nucleases (correspondente ao controlo negativo), ou 1µl de cDNA (produto resultante da reação de RT descrita em 2.3) ou de uma diluição de 1/10 de uma solução contendo o plasmídeo onde as sequências nucleotídicas otimizadas foram originalmente clonadas (correspondente ao controlo positivo), perfazendo um volume final de 25µl. A reação de PCR foi realizada no termociclador *Hybaid PCR Express* e as suas condições encontram-se descritas na Tabela 2.6.

Os produtos amplificados foram analisados por eletroforese em gel de agarose a 1,5%, tal como descrito em 2.6. Para a análise foi aplicado, em cada um dos poços do gel de agarose, 5µl de amostra de cada um dos produtos que resultou das reações de PCR.

**Tabela 2.6.** Condições de reação utilizadas na amplificação, por PCR, das regiões codificantes de C e NS3hel para análise de moléculas recombinantes de DNA plasmídico.

Etapa	Temperatura (°C)	Tempo de duração	Número de ciclos
<b>Desnaturação inicial</b>	95	2 min	1
<b>Desnaturação</b>	95	30 seg	35
<b>Hibridação</b>	65	30 seg	
<b>Extensão</b>	72	45 seg	
<b>Extensão final</b>	72	5 min	1
<b>Manutenção</b>	4	∞	1

### 2.8.7. Sequenciação de DNA

Para a sequenciação das moléculas recombinantes de DNA plasmídico foi efetuada, primeiramente, a purificação das mesmas (obtidas em 2.8.5), utilizando o sistema comercial *DNA Clean & Concentrator-5<sup>TM</sup>*, tal como descrito em 2.7. O rendimento do processo de purificação foi avaliado através da análise, por eletroforese em gel de agarose a 1%, tal como descrito em 2.6.

As moléculas recombinantes de DNA plasmídico purificadas foram enviadas para a empresa STAB VIDA (Portugal) no intuito de se proceder à sua sequenciação. O processo recorreu à técnica de terminação da polimerização em cadeia com didesoxi-ribonucleótidos, descrita inicialmente por Sanger (Sanger *et al.*, 1977). As moléculas de DNA recombinante foram sequenciadas parcialmente (região próxima do local de ligação do inserto ao vetor) e em ambos os sentidos, utilizando *primers* universais, complementares à sequência do vetor.

## 2.9. Expressão de proteínas recombinantes de ZIKV

Para a expressão de proteínas recombinantes de ZIKV foram escolhidas como estirpes de *E. coli*, *BL21(DE3)pLysS* [ $F^- ompT hsdS_B (r_B^- m_B^-)$  gal dcm (DE3) pLysS ( $Cm^R$ )] e *Rosetta(DE3)pLysS* [ $F^- ompT hsdS_B (r_B^- m_B^-)$  gal dcm (DE3) pLysSRARE ( $Cm^R$ )] (Novagen, Alemanha).

A preparação de células competentes de *E. coli* *BL21(DE3)pLysS* e *Rosetta(DE3)pLysS* e sua transformação foram efetuadas tal como descrito em 2.8.3 e

2.8.4, respetivamente. No entanto, no passo de preparação de células competentes, o meio LB líquido foi suplementado com cloranfenicol (Cam; 35µg/ml). Para a transformação das células competentes foram adicionados 12µl de solução contendo o DNA plasmídico extraído (ver 2.8.5). As suspensões bacterianas foram semeadas em placas de Petri com meio LB sólido suplementado com Kan (30µg/ml) e Cam (35µg/ml).

Os clones bacterianos contendo plasmídeos recombinantes foram reinoculados e a 1ml de cada cultura bacteriana adicionou-se 500µl de glicerol a 50% estéril, para posterior conservação a -80°C.

### **2.9.1. Teste de expressão proteica**

O procedimento foi efetuado com base nos protocolos ligeiramente modificados descritos em *pET System Manual 11<sup>th</sup> Edition* (Novagen, Alemanha). Colónias bacterianas isoladas (obtidas em 2.9) foram repicadas, com um palito estéril, para tubos com 3ml de meio LB líquido suplementado com Kan (30µg/ml) e Cam (35µg/ml), e incubadas durante a noite a 37°C, com agitação (220rpm), até à obtenção de uma cultura bacteriana saturada. Após esse período, 150µl de cultura saturada obtida foram diluídos em 15ml (diluição 1:100) de meio LB líquido suplementado com Kan (30µg/ml) e Cam (35µg/ml), ao que se seguiu uma incubação nas condições de crescimento celular descritas acima até ter sido atingida uma D.O.<sub>600nm</sub> de ≈0,8. Seguidamente, retirou-se 3ml de cultura para um tubo e adicionou-se isopropil-β-D-tiogalactopiranosídeo (IPTG; 1mM final), induzindo a expressão proteica durante 3h a 37°C com agitação (200-220rpm). A um outro tubo foi adicionado apenas 3ml de cultura, correspondendo à cultura não induzida. Finda a indução foram recolhidas alíquotas de 1ml das culturas bacterianas, em condições assépticas. As células bacterianas foram colhidas por centrifugação a 16000g durante 1min, à temperatura ambiente. Após decantação do sobrenadante, o sedimento celular foi ressuspensionado em 100µl de solução tampão de lise fria (20mM Tris-HCl pH 7,4; 0,5M NaCl; solução base: alteração da sua composição ao longo do projeto mantendo a mesma designação) e 100µl de solução tampão de amostra (Anexo II), para posterior avaliação do nível de expressão das proteínas por eletroforese em gel de poliacrilamida na presença de dodecilsulfato de sódio (SDS-PAGE) ou tricina-SDS-PAGE.

### **2.9.2. Testes de otimização de produção de proteínas recombinantes**

No presente estudo foram avaliados vários parâmetros de forma a otimizar a produção de proteínas recombinantes. O primeiro parâmetro avaliado foi a temperatura de crescimento, tendo sido testadas três temperaturas diferentes: 16, 25 e 37°C. A indução foi efetuada como descrito em 2.9.1. As culturas bacterianas foram induzidas durante a noite com agitação (170-180rpm), numa câmara frigorífica (16°C) ou num agitador orbital (25°C). Para avaliar o eventual efeito de variação da concentração de indutor (IPTG) foram testadas duas condições, correspondentes à utilização de concentrações finais de IPTG de 0,5 e 1mM. A indução foi efetuada como descrito em 2.9.1., a 16°C ou 25°C com 0,5mM, durante a noite com agitação (170-180rpm) e a 25°C com 0,5mM, durante 4h e agitação regular (200-220rpm). Um outro parâmetro avaliado foi a heterogeneidade da população bacteriana. Assim teve-se em conta a escolha de diferentes colónias bacterianas isoladas (obtidas em 2.9). A indução foi efetuada como descrito em 2.9.1. Para a otimização de produção das proteínas também foi efetuado um ensaio de cinética de produção, no qual a indução foi efetuada como descrito em 2.9.1. No entanto, após a adição de IPTG procedeu-se à recolha de alíquotas de 1ml das culturas bacterianas em diferentes momentos: 1h, 2h, 3h e 5h. Todas as condições foram avaliadas ao nível de expressão das proteínas por tricina-SDS-PAGE.

### **2.9.3. Avaliação da toxicidade proteica através da monitorização do crescimento bacteriano em meio líquido**

Para avaliação da toxicidade decorrente da expressão das proteínas heterólogas desejadas procedeu-se à indução como descrito em 2.9.1. No entanto foi adicionado IPTG (1mM final) quando atingida uma D.O.<sub>600nm</sub>, de  $\approx 0,4$ . No momento do começo da indução foi iniciado o registo da D.O.<sub>600nm</sub> das diferentes culturas de 30 em 30min.

### **2.9.4. Testes de otimização de solubilidade proteica**

A verificação da solubilidade das proteínas recombinantes foi iniciada através da inoculação de suspensões celulares correspondentes a clones bacterianos contendo plasmídeos recombinantes (culturas puras e concentradas), conservadas a 4°C em 3ml de meio LB líquido suplementado com Kan (30µg/ml) e Cam (35µg/ml), seguindo uma incubação a 37°C com agitação (220rpm) durante a noite. Após esse período, 150µl de



cultura saturada obtida foram diluídos em 15ml (diluuição 1:100) de meio LB líquido suplementado com Kan (30µg/ml) e Cam (35µg/ml), ao que se seguiu uma incubação nas condições de crescimento celular descritas acima até ter sido atingida uma D.O.<sub>600nm</sub>, de  $\approx 0,8$ . De seguida, às culturas bacterianas foi adicionado IPTG (1mM final) de forma a induzir a expressão proteica a partir das variantes de pET-29a durante 3h, a 37°C, com agitação (200-220rpm). As células bacterianas foram colhidas por centrifugação a 3220g, durante 15min, a 4°C. Após decantação do sobrenadante, o sedimento celular foi conservado a -20°C até ser utilizado, ou ressuscitado em 1,5ml (fator de concentração: 10X) de solução tampão de lise fria (20mM Tris-HCl pH 7,4; 0,5M NaCl; 0,1% (v/v) Triton X-100; 10% (v/v) glicerol). As suspensões celulares foram lisadas por submissão a 5 ciclos de congelação/descongelação em azoto líquido/37°C seguidos de sonicação (*Ultrasonic processor UP200S* - Hielscher, Alemanha) usando ciclos com 80% de amplitude e 5 pulsos de 30seg intervalados com 30seg de pausa, mantendo sempre os tubos em gelo. Os lisados foram submetidos a uma centrifugação a 14000g durante 10min a 4°C, obtendo-se duas frações: solúvel (sobrenadante) e insolúvel (sedimento). O volume de sobrenadante foi transferido para um novo microtubo. A fração insolúvel foi lavada com 750µl de solução tampão de lise fria, seguida de centrifugação a 10000g durante 5min a 4°C. O processo de lavagem foi repetido e, no final, o sedimento foi ressuscitado em 1,5ml de solução de SDS a 1% (m/v). Para novos microtubos foram transferidos 100µl de cada uma das frações (solúvel e insolúvel) e 100µl de solução tampão de amostra, para posterior avaliação do nível de expressão das proteínas (conservado a -20°C). O restante volume foi guardado a -20°C.

No presente estudo foram avaliados vários parâmetros de forma a melhorar a solubilidade proteica.

O primeiro parâmetro avaliado foi a temperatura de crescimento, tendo sido testadas três temperaturas diferentes: 16, 25 e 37°C. A indução foi efetuada como descrito acima. As culturas bacterianas foram induzidas durante a noite com agitação (170-180rpm), numa câmara frigorífica (16°C) ou num agitador orbital (25°C). A 25°C procedeu-se também a induções durante 1h, 3h e 6h. Quando a indução decorreu a 37°C, além das 3h de indução, também se realizou uma indução de 1h.

Um outro parâmetro avaliado foi a fase de crescimento na qual a indução da expressão proteica foi efetuada. A indução foi realizada como descrito acima e foram testadas as

D.O.<sub>600nm</sub>, de  $\approx 0,4$ ,  $0,8$  e  $1,2$ . A D.O.<sub>600nm</sub>, de  $\approx 0,4$  e  $1,2$ , as culturas bacterianas foram induzidas durante 2h.

Para a otimização de solubilidade proteica foram efetuados ensaios no qual foi adicionado ao meio LB líquido o aditivo D-sorbitol (VWR International, E.U.A). As concentrações testadas foram  $0,5M$ ,  $0,75M$  e  $1M$ . A indução foi realizada como descrito acima, no entanto, a cultura saturada obtida, após inoculação durante a noite foi diluída em meio LB líquido suplementado com D-sorbitol,  $0,1M$  de solução tampão fosfato de sódio ( $77,4mM Na_2HPO_4$ ;  $22,6mM NaH_2PO_4 \cdot H_2O$ ), Kan ( $30\mu g/ml$ ) e Cam ( $35\mu g/ml$ ). Quando atingida uma D.O.<sub>600nm</sub>, de  $\approx 0,8$  adicionou-se, primeiramente,  $80mM$  de solução cloreto de potássio (KCl) e depois IPTG para uma concentração final de  $1mM$ .

Para avaliar o eventual impacto de lise celular na otimização de solubilidade proteica foram utilizados dois mecanimos de lise. Neste caso foram testados 5 ciclos de congelamento/descongelamento em azoto líquido e a sonicação. Foram testadas duas condições, o conjunto 5 ciclos de congelamento/descongelamento em azoto líquido/ $37^\circ C$  seguido de sonicação, e 5 ciclos de congelamento/descongelamento em azoto líquido/ $37^\circ C$  apenas. A indução foi efetuada como descrito acima.

Todas as condições foram avaliadas ao nível de expressão das proteínas por SDS-PAGE.

#### **2.9.5. Avaliação do nível de expressão proteica por SDS-PAGE**

Para avaliação do nível de expressão das proteínas, os extratos proteicos obtidos em 2.9.1, 2.9.2 e 2.9.4 foram fervidos durante 10min a  $100^\circ C$  e, posteriormente, analisados por eletroforese em tricina-SDS-PAGE ou SDS-PAGE. Inicialmente foi realizado um método modificado de tricina-SDS-PAGE, descrito por Haider *et al.*, 2010. Também foi realizado o método convencional SDS-PAGE (Laemmli, 1970). Os géis de resolução (Anexo II) continham uma concentração de 12% e 13% de poliacrilamida, respetivamente e os géis de concentração (Anexo II) continham uma concentração de 4% e 5% de poliacrilamida, respetivamente. Para a migração eletroforética no gel de concentração foi aplicada uma voltagem constante de 50V e no gel de resolução aumentou-se para 100V. A massa molecular das proteínas foi estimada a partir da migração relativa de um marcador de proteínas com massas moleculares conhecidas (*Precision Plus Protein™ Dual Color Standards* – Bio-Rad, E.U.A.). Finda a eletroforese, os géis foram corados

com uma solução comercial, *BlueSafe* (NZYtech, Portugal), por incubação com agitação orbital suave, à temperatura ambiente, até as bandas proteicas estarem claramente visíveis.

#### **2.9.6. Recuperação da estrutura tridimensional de proteínas desnaturadas (*Refolding* proteico)**

Para o *refolding* proteico procedeu-se à indução da produção proteica como descrito em 2.9.4. No entanto ressuspendeu-se o sedimento celular (fração insolúvel) em 1,5ml de solução tampão de lise com ureia (8M final) (Anexo II). De seguida fez-se um conjunto de diálises sucessivas, com recurso a três tampões com concentrações decrescentes de ureia (1º: 50mM Tris-HCl pH 7,4; 0,5M NaCl; 10% (v/v) glicerol; 4M ureia; 2º: 50mM Tris-HCl pH 7,4; 0,5M NaCl; 10% (v/v) glicerol; 2M ureia; 3º: 50mM Tris-HCl pH 7,4; 0,5M NaCl; 5% (v/v) glicerol). As duas primeiras diálises foram realizadas durante 2h30min e a terceira durante a noite, a 4°C, com agitação suave. O produto resultante foi submetido a uma centrifugação a 14000g durante 10min a 4°C, obtendo-se duas frações: solúvel (sobrenadante) e insolúvel (sedimento). O volume de sobrenadante foi transferido para um novo microtubo e o sedimento foi ressuspendido em 1,5ml de solução de SDS a 1% (m/v). Para novos microtubos foram transferidos 100µl de cada uma das frações (solúvel e insolúvel) e 100µl de solução tampão de amostra (conservados a -20°C). O restante volume foi guardado a -20°C.

A avaliação do sucesso de *refolding* proteico foi efetuada por SDS-PAGE, tal como descrito em 2.9.5.

#### **2.9.7. Verificação da presença de cauda de His<sub>6</sub>, por purificação em condições desnaturantes, nas proteínas recombinantes produzidas**

A verificação da presença de caudas de His<sub>6</sub> nas proteínas recombinantes produzidas por purificação em condições desnaturantes foi efetuada utilizando o sistema comercial *Ni-NTA Spin Kit* (QIAGEN, Alemanha) e seguindo as instruções recomendadas pelo fabricante. O sedimento celular resultante de 15ml de cultura bacteriana induzida (tal como descrito em 2.9.4) foi ressuspendido em 700µl de solução tampão de lise (Anexo II). De seguida, o lisado celular foi incubado com agitação suave à temperatura ambiente durante 15min, e centrifugado a 12000g, à temperatura ambiente durante 25min. Para um

novo microtubo foram transferidos 20µl do sobrenadante obtido e 20µl de solução tampão de amostra (conservado a -20°C). O restante volume foi aplicado numa coluna com uma resina com íões níquel, de forma a permitir a fixação da proteína à coluna. De seguida foi adicionada solução tampão de lavagem (Anexo II) à coluna para remoção de contaminantes/ligações não específicas. A proteína foi eluída por adição de 200µl de solução tampão de eluição (Anexo II) à coluna. De seguida fez-se a recolha do eluato e repetiu-se este passo para permitir uma eluição completa. Para novos microtubos foram transferidos 20µl de cada filtrado e eluato e 20µl de tampão de amostra (conservado a -20°C).

Em alternativa ao descrito anteriormente, às soluções tampão de lise, lavagem e eluição foram adicionados 0,5M de NaCl. O sedimento celular resultante de 15ml de cultura bacteriana induzida (tal como descrito em 2.9.4) foi ressuspensionado em 700µl de solução tampão de lise (Anexo II) com 0,5M NaCl e 35µl de solução de inibidores de proteases (1X final) (*cOmplete™, EDTA-free Protease Inhibitor Cocktail* – Roche, Suíça). O restante procedimento decorreu tal como descrito em 2.9.4. No entanto, o sobrenadante resultante da centrifugação que permite obter as frações solúvel e insolúvel foi aplicado numa coluna com uma resina com íões níquel. O processo de purificação foi efetuado com descrito acima, no entanto foram utilizadas as soluções tampão de lavagem e eluição com 0,5M NaCl.

A avaliação do processo de purificação em condições desnaturantes foi efetuada por SDS-PAGE, tal como descrito em 2.9.5.

### **2.9.8. Purificação de proteínas em condições nativas**

A purificação em condições nativas de proteínas recombinantes com caudas de His<sub>6</sub> foi efetuada utilizando o sistema comercial *Ni-NTA Spin Kit* e seguindo as instruções recomendadas pelo fabricante. O sedimento celular resultante de 15ml de cultura bacteriana induzida (tal como descrito em 2.9.4) foi ressuspensionado em 1,4ml de solução tampão de lise (Anexo II). O restante procedimento realizou-se tal como descrito em 2.9.4, no entanto, o sobrenadante resultante da centrifugação que permite obter as frações solúvel e insolúvel foi aplicado numa coluna com uma resina com íões níquel. O processo de purificação foi efetuado com descrito em 2.9.7, no entanto com a utilização de soluções

tampão de lavagem e eluição com componentes diferentes (Anexo II). A eluição foi realizada pela adição à coluna de 0,5M de solução de imidazole.

A avaliação do processo de purificação em condições nativas foi efetuada por SDS-PAGE, tal como descrito em 2.9.5.

#### **2.9.9. Verificação da presença de cauda de His<sub>6</sub>, por *western blot*, nas proteínas recombinantes produzidas**

O método *western blot* também permitiu a verificação da presença de caudas de His<sub>6</sub>, desta feita utilizando anticorpos monoclonais anti-His<sub>6</sub>. O procedimento de detecção dos complexos antígeno/anticorpo primário/anticorpo secundário conjugado foi efetuado utilizando o sistema comercial *Amersham<sup>TM</sup> ECL Western Blotting Detection Kit* (GE Healthcare Life Sciences, E.U.A., ver secção 2.9.9.3) e seguindo as instruções recomendadas pelo fabricante. O *western blot* foi iniciado pela separação das proteínas por SDS-PAGE, como descrito em 2.9.5, e a sua transferência para um suporte sólido (membrana), como descrito abaixo.

##### **2.9.9.1. Transferência de proteínas para uma membrana**

A transferência das proteínas do gel de poliacrilamida para uma membrana de fluoreto de polivinilideno (PVDF) (*Amersham<sup>TM</sup> Hybond P* - GE Healthcare Life Sciences, E.U.A.) foi realizada utilizando o sistema semi-seco *Trans-Blot SD Semi-Dry Transfer Cell* (Bio-Rad, E.U.A.), seguindo as instruções recomendadas pelo fabricante. A membrana hidratada em solução tampão *Towbin* (Anexo II) juntamente com o gel de poliacrilamida foram colocados entre folhas de papel *Whatman 3MM*, igualmente embebidas em solução tampão *Towbin*. A membrana ficou junto ao ânodo e o gel ao cátodo. A transferência decorreu durante 1h a 15V.

##### **2.9.9.2. Bloqueio da membrana e incubação com anticorpos**

A membrana de PVDF foi incubada com solução tampão de bloqueio (Anexo II) contendo 2% (m/v) de *ECL<sup>TM</sup> Advance Blocking Agent* durante 1h30min à temperatura ambiente com agitação suave. Após a incubação, a membrana foi enxaguada duas vezes com solução tampão TBST (Anexo II). De seguida, a membrana foi incubada com

anticorpo primário, *i.e.*, anticorpo monoclonal de rato anti-histidina (GE Healthcare Life Sciences, E.U.A.), diluído em solução tampão de bloqueio (1:10.000), durante 1h à temperatura ambiente com agitação suave. A solução contendo o anticorpo primário foi decantada e a membrana foi enxaguada duas vezes com solução tampão TBST. De seguida foi realizada uma lavagem de 15min, seguida de três outras de 5min, também usando solução tampão TBST. A membrana foi incubada com anticorpo secundário, *i.e.*, anticorpo anti-IgG de rato, produzido em coelho, conjugado com peroxidase (Sigma-Aldrich, E.U.A.), diluído em solução tampão de bloqueio (1:10.000), durante 1h à temperatura ambiente com agitação suave. A solução contendo o anticorpo secundário foi decantada e a membrana foi enxaguada duas vezes com solução tampão TBST. De seguida foi realizada uma lavagem de 15min e três de 5min, também em solução tampão TBST.

#### **2.9.9.3. Detecção de proteínas por quimioluminescência**

Para a detecção dos complexos formados entre proteínas com caudas de histidina e os anticorpos utilizados na sua detecção, primeiramente foi retirado o excesso de solução tampão TBST e de seguida, a membrana foi incubada com uma solução de detecção (Anexo II) durante 5min à temperatura ambiente. Após o excesso de solução de detecção ter sido retirado, a membrana foi recoberta com película aderente e, de seguida colocada com face contendo os complexos antígeno/anticorpos virados para cima, numa cassete de autorradiografia. Em seguida, sobre a membrana foi colocado filme autorradiográfico (*Amersham<sup>TM</sup> Hyperfilm ECL* - GE Healthcare Life Sciences, E.U.A.), a cassete fechada e exposição permitida durante aproximadamente 10 segundos. O filme autorradiográfico foi revelado, e repetiu-se o processo de exposição até a obtenção de um sinal com menor ruído possível.

### **3. RESULTADOS E DISCUSSÃO**

#### **3.1. Otimização das sequências nucleotídicas codificantes de C e NS3hel para ajustamento da preferência de utilização de códons em *E. coli***

Atualmente, ZIKV surge como um agente patogénico humano altamente relevante, não só pela possibilidade que demonstrou de estar na origem de surtos de grandes dimensões, em regiões do globo com populações ativas de mosquitos do género *Aedes*, como pela sua associação àquilo que se veio a designar como síndrome congénita a ZIKV e à síndrome de Guillain-Barré. Contudo, até ao presente, pouco se conhece sobre a estratégia replicativa utilizada pelo ZIKV nas células que lhe são permissivas, incluindo que tipo de interações são estabelecidas com a maquinaria celular durante o ciclo replicativo, ou de que maneira a fisiologia celular e a presença de outros vírus podem influenciar estas interações. A resposta a estas questões assenta, em parte, no desenvolvimento de ferramentas que permitam estes, e outros, estudos e que incluem, como exemplos, a construção de mutantes virais, anticorpos específicos, ou proteínas virais recombinantes a partir das quais possa ser feita a determinação da sua estrutura tridimensional após cristalização. Adicionalmente, para o caso das proteínas virais que apresentem atividade enzimática (em particular NS3 e NS5), estas proteínas recombinantes podem ainda ser utilizadas em ensaios de caracterização da sua atividade enzimática, visando o rastreio, *in vitro*, de compostos que as possam especificamente inibir, e que possam ser explorados no sentido da produção de fármacos antivirais.

Assim, tendo em conta a disponibilidade relativa no que diz respeito a ferramentas básicas que possam ser usadas para melhor compreender a biologia de ZIKV foi estabelecido como objetivo principal deste projeto de tese, a otimização da produção/purificação de antígenos virais de ZIKV em *E. coli*. Como objeto deste estudo foram escolhidos dois antígenos com funções biológicas e características distintas: por um lado a proteína estrutural da cápside (a proteína C), e um muito provável constituinte do complexo replicativo (aqui designado NS3hel), e que corresponde à região C-terminal da proteína NS3 nativa, à qual está associada atividade de helicase de RNA.

Dado que estes objetivos assentam na produção de proteínas virais em quantidade e com elevado grau de pureza e solubilidade, *E. coli* surge como um dos possíveis sistemas onde a sua produção pode ser executada. Ainda que as proteínas de interesse sejam

codificadas por um vírus que infeta células eucariontes, *E. coli* apresenta variadas vantagens relativamente aos sistemas baseados em células eucariontes, nomeadamente, uma fácil sub-cultura e conservação, rápido crescimento em meios de composição simples e pouco dispendiosa, fácil manipulação genética e existência de um vasto arsenal de estirpes nas quais a expressão proteica pretendida pode ser otimizada e da qual, normalmente, é obtido um elevado rendimento a um custo relativamente baixo. Todas estas características fazem de *E. coli* o hospedeiro de eleição para a produção de proteínas recombinantes de origem heteróloga (Yin *et al.*, 2007; Demain e Vaishnav, 2009; Rosano e Ceccarelli, 2014). Contudo, ainda assim, esta bactéria pode levar a uma produção limitada das proteínas de interesse devido a diversos fatores. São destes exemplos, a impossibilidade de ocorrência de modificações pós-traducionais nas proteínas a serem produzidas, o desajuste na utilização de codões das sequências que as codificam, e a produção das proteínas desejadas com elevado grau de insolubilidade e/ou com conformação incorreta.

Considerando que as proteínas estudadas no decurso deste trabalho não sofrem modificações pós-traducionais no sistema em que naturalmente são expressas, um dos fatores que, *a priori*, mais poderia condicionar a sua produção em quantidade usando *E. coli* como hospedeiro de expressão seria o desajuste na utilização de codões que codificam as proteínas de ZIKV e os que normalmente integram as sequências codificantes de *E. coli*. Este desajuste ocorre quando a frequência de utilização de codões sinónimos na sequência nucleotídica que codifica as proteínas de interesse difere da do hospedeiro em que esta se virá a expressar de forma heteróloga. Tal situação decorre do facto de, em cada célula, a população de tRNAs refletir de forma aproximada, a frequência de utilização da população de codões que integram as sequências codificantes desse contexto celular. Portanto, se o mRNA de interesse contém codões que raramente são encontrados nos genes de *E. coli*, aquando da síntese das proteínas recombinantes, uma baixa disponibilidade de alguns tRNAs pode levar à incorporação de aminoácidos erróneos e/ou ao truncamento do polipéptido desejado (como resultado de uma paragem na síntese proteica), afetando os níveis de expressão e/ou a qualidade conformacional das proteínas recombinantes produzidas (Dong *et al.*, 1996; Gustafsson *et al.*, 2004; Terpe, 2006; Rosano e Ceccarelli, 2014). Assim, tendo em conta esse problema procedeu-se à análise nas sequências nucleotídicas que codificam C e NS3hel no genoma de ZIKV



(número de acesso LC002520), procurando identificar a presença de codões que apenas raramente são encontrados em *E. coli*.

De facto, e tal como seria expectável tratando-se ZIKV de um vírus de células eucariontes, verificou-se a presença de codões raros nas sequências analisadas. Em *E. coli* encontram-se definidos como codões raros, os codões AGA (arginina, Arg), AGG (Arg), CGA (Arg), CGG (Arg), GGA (glicina, Gly), AUA (isoleucina, Ile), CUA (leucina, Leu) e CCC (prolina, Pro) (Novy *et al.*, 2001). Na sequência codificante da proteína C verificou-se a presença de 24 codões raros num total de 122 codões ( $\approx 20\%$ ), enquanto que na sequência codificante de NS3hel foram identificados 21 codões raros num total de 117 codões ( $\approx 18\%$ ) (Tabela 3.1). Considerando estes resultados definiu-se como uma etapa prioritária neste projeto, a otimização das regiões codificantes dos polipéptidos C e NS3hel, a qual permitiria a geração de novas sequências codificantes com codões reconhecidos pelo conjunto de tRNAs de *E. coli*, sem modificar a sequência primária das proteínas pretendidas. O processo de otimização foi obtido através da síntese artificial (NZYtech, Portugal) das sequências codificantes pretendidas, tendo em conta fatores, tais como, a adaptabilidade dos codões ao hospedeiro a ser utilizado, a estrutura do mRNA a ser produzido e a eventual presença de elementos reguladores da expressão génica na região codificante a ser utilizada.

**Tabela 3.1.** Frequência absoluta de codões raros em *E. coli* nas regiões codificantes dos antígenos C e NS3hel, na sequência correspondente ao genoma completo de ZIKV (número de acesso LC002520).

<b>Proteínas</b> <b>Codões raros</b>	<b>C</b>	<b>NS3hel</b>
AGG (Arg)	3	4
AGA (Arg)	5	4
CGG (Arg)	1	2
CGA (Arg)	0	0
GGA (Gly)	7	4
AUA (Ile)	4	5
CUA (Leu)	2	1
CCC (Pro)	2	1
	Total=24	Total=21

Uma vez que a preferência de utilização da população de codões reflete de uma forma genérica a abundância da população de tRNAs na célula, a otimização do sistema de tradução pelo aumento da disponibilidade de tRNAs, normalmente de baixa abundância e, portanto, limitante, pode ser conseguida de forma alternativa, suplementando a sua

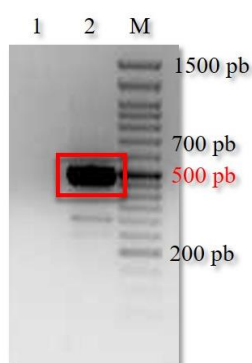
presença, *in trans*, usando plasmídeos que transportam os genes que os codificam (ver secção 3.4.).

### 3.2. Amplificação de sequências codificantes de ZIKV

#### 3.2.1. Confirmação, por RT-PCR, da presença de genoma de ZIKV numa suspensão viral

A confirmação da presença de genomas de ZIKV num extrato de RNA total, preparado a partir de uma suspensão viral e conservada a  $-80^{\circ}\text{C}$ , foi efetuada através da amplificação, por RT-PCR, de uma sequência genómica correspondendo a parte da região que codifica parte dos componentes estruturais do vírus (coordenadas 165-661 na sequência de genoma completo, com o número de acesso LC002520). Esta pressupõe uma reação de RT visando a síntese de cDNA a partir de RNA total extraído do sobrenadante de cultura de células Vero E6 infetadas por ZIKV (estirpe MR766; ver secções 2.2 e 2.3), a qual foi seguida por uma amplificação por PCR (ver secção 2.5.1).

A secção do genoma escolhida como alvo desta amplificação baseou-se na pesquisa de regiões conservadas entre as várias sequências de referência alinhadas, compreendendo parte da sequência que codifica para as proteínas C ( $\approx 310\text{nt}$ ) e prM ( $\approx 190\text{nt}$ ). A análise do produto obtido como resultado da amplificação executada revelou a presença de um fragmento de DNA com tamanho de  $\approx 500$  pares de bases (pb) (Figura 3.1).



**Figura 3.1.** Confirmação da presença de genomas de ZIKV num extrato de RNA total, por amplificação por RT-PCR de um fragmento de  $\approx 500\text{pb}$  (após separação por eletroforese em gel de agarose a 1,5%). 1 - controlo negativo; 2 – produto de amplificação desejado (ver texto); M – marcador de massa molecular (NZYDNA Ladder VI).

### 3.2.2. Amplificação das regiões codificantes de C e NS3hel a partir de sequências de genótipo selvagem e versões sintéticas com sequências codificantes otimizadas

As sequências nucleotídicas de genótipo selvagem de ZIKV e as versões sintéticas com códons otimizados que codificam C e NS3hel foram amplificadas por PCR, a partir do cDNA resultante da reação de RT (ver secção 2.3) ou a partir de plasmídeos em que foram inicialmente fornecidas, respetivamente (ver secção 2.1).

Numa primeira abordagem, os pares de *primers* complementares às sequências codificantes de C e NS3hel foram desenhados de forma a permitir a amplificação da totalidade das mesmas sob a forma de fragmentos de 366pb e 351pb, respetivamente. Contudo, após os primeiros testes de expressão proteica (ver secções seguintes), quando esta era induzida a partir de uma cultura em que o vetor de expressão comportava um inserto correspondente à totalidade da sequência codificante de C era exercida uma elevada toxicidade sobre o hospedeiro de expressão, que afetava visivelmente a viabilidade celular, e que se traduzia na lise, quase imediata da cultura em questão após a adição do agente indutor (IPTG) (ver secção 3.4.1). De forma a corrigir este problema, a análise da sequência primária de C permitiu confirmar a existência de uma região localizada no extremo carboxilo (C-terminal) que deverá corresponder, por alinhamento com outras regiões homólogas das proteínas da cápside de outros flavivírus, a uma secção transmembranar (hidrofóbica). A presença destes domínios transmembranares em proteínas sobre-expressas está frequentemente associada a fenómenos de toxicidade, possivelmente pela destabilização de membranas e/ou pela formação de poros, podendo levar à morte do organismo de expressão heteróloga (Montigny *et al.*, 2004). Na proteína C, esta região hidrofóbica serve como péptido sinal na translocação de prM para o RE durante a tradução da poliproteína viral. Após translocação de prM, a secção transmembranar é clivada, tornando C uma proteína madura (Lindenbach *et al.*, 2013; Barzon *et al.*, 2016b). Tendo isto em conta, a região hidrofóbica não se encontra presente na proteína que se pretende produzir, por isso, optou-se por redesenhar um dos *primers* de amplificação (*primer R*) e assim, eliminar a região codificante deste segmento transmembranar em C-terminal da variante de C de ZIKV a ser expressa no decurso deste trabalho.

Na respetiva extremidade 5' de cada um dos *primers* utilizados nas amplificações pretendidas foram adicionados nucleótidos não complementares às sequências nucleotídicas a amplificar, e que corresponderam a locais de clivagem por endonucleases de restrição do tipo II. As sequências de restrição introduzidas eram reconhecidas pelas enzimas *NotI* e *NdeI*. A escolha das mesmas teve em conta a ausência da sua sequência de restrição nas sequências nucleotídicas que codificam C e NS3hel, bem como a disponibilidade destes locais de restrição enzimática nas sequências do vetor de expressão utilizado (pET-29a). Adicionalmente, estas duas enzimas apresentam atividade máxima na mesma solução tampão, permitindo a execução de uma única reação de digestão enzimática em que ambas são utilizadas em conjunto.

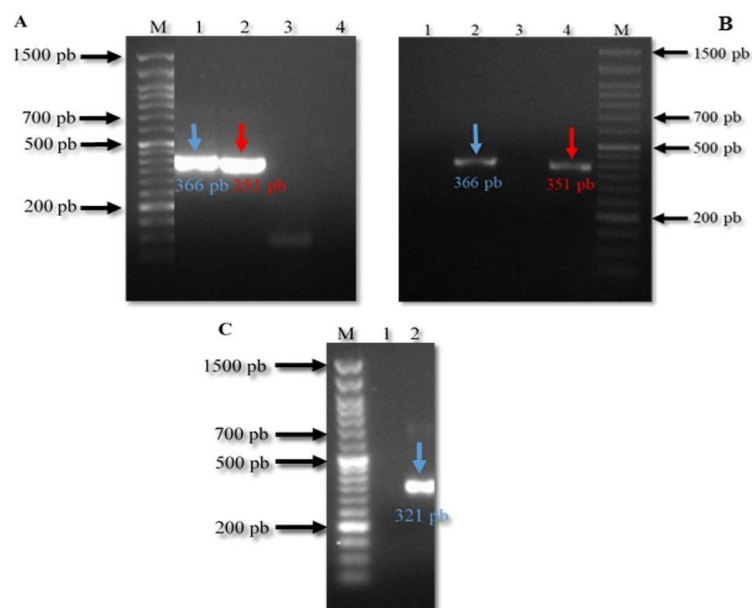
No caso de NS3hel, a sequência de restrição de *NdeI* (CATATG) foi utilizada para introduzir o codão de iniciação (ATG), correspondente ao resíduo metionina, uma vez que esta proteína é naturalmente traduzida no contexto de uma poliproteína e o seu extremo N-terminal resulta de uma clivagem proteolítica. Contrariamente, a proteína C é a primeira proteína na poliproteína viral a ser traduzida, pelo que contém ATG como primeiro codão. Neste caso foram apenas adicionados ao *primer* de amplificação, os nucleótidos necessários para completar a sequência de restrição de *NdeI*. A clivagem pela enzima de restrição *NotI* da extremidade 3' dos fragmentos amplificados e de pET-29a, seguida da ligação entre os mesmos, possibilitou a adição às sequências codificantes das proteínas a expressar uma cauda de hexa-histidina (His<sub>6</sub>; ver secção seguinte), permitindo a sua subsequente purificação por cromatografia de afinidade. Não obstante, entre a sequência codificante das proteínas virais e da sequência His<sub>6</sub>, as manipulações genéticas executadas resultaram na manutenção de parte da sequência do vetor, que codifica um péptido de ligação de 5 aminoácidos entre C/NS3hel e His<sub>6</sub>. No entanto, não se antecipa que estes aminoácidos possam interferir no processo de expressão heteróloga e subsequente purificação proteica.

A sequência de restrição da enzima *NdeI* foi adicionada aos *primers F* e a de *NotI* aos *primers R*. Para além da sequência de restrição foram também adicionados nucleótidos não complementares ao extremo 5' de cada um dos *primers* com o intuito de melhorar a eficácia de corte das enzimas de restrição. A estes últimos nucleótidos designou-se como sequência espaçadora e à sequência de restrição de *NdeI* foram adicionados 9nt e à de *NotI*, 8nt.

O protocolo de amplificação das sequências nucleotídicas que codificam C e NS3hel de genótipo selvagem de ZIKV e versões sintéticas com codões otimizados (ver secções 2.5.2 e 2.5.3) foi elaborado de forma a aumentar a especificidade do processo de PCR, pelo que foi tido em conta fatores como, a temperatura de hibridação dos *primers* e a DNA polimerase utilizada. Normalmente, num protocolo de PCR, a temperatura de hibridação dos *primers* utilizada está cerca de 5°C abaixo do T<sub>m</sub> (temperatura de dissociação) calculado. Contudo, neste protocolo, definiu-se o valor de 65°C, o qual é superior, em 6°C, ao T<sub>m</sub> teórico calculado *in silico* (*OligoAnalyzer 3.1*, disponível em <https://eu.idtdna.com/calc/analyzer>) de forma a aumentar a restringência e, consequentemente, a especificidade das amplificações.

A amplificação foi efetuada através da utilização do sistema comercial *Supreme NZYTaQ DNA polymerase*. Neste sistema, pode ser encontrada uma versão modificada da Taq polimerase, a qual se encontra inativa à temperatura ambiente. A enzima apenas apresenta atividade enzimática funcional quando incubada a uma temperatura elevada (95°C, este processo é designado *hot-start PCR*), permitindo assim aumentar a especificidade das amplificações pretendidas. Ainda que, idealmente, devesse ter sido utilizada uma polimerase com atividade corretora de provas (*proof-reading*) optou-se pela sua não utilização, pois os fragmentos de DNA a amplificar eram de reduzido tamanho, pelo que a probabilidade de inserção de erros pela polimerase utilizada seria muito baixa. Por outro lado, todas as sequências nucleotídicas clonadas vieram a ser confirmadas por sequenciação, o que permitiria, se bem que *a posteriori*, a deteção de eventuais mutações não-desejadas nos clones gerados.

Numa primeira abordagem, a análise dos produtos obtidos como resultado das amplificações das sequências nucleotídicas de genótipo selvagem de ZIKV e das versões sintéticas com codões otimizados revelou a presença de fragmentos de DNA (Figura 3.2 A e B), com tamanho aproximado de 370pb (C, tamanho esperado=366pb) e 350pb (NS3hel, tamanho esperado=351pb). A análise do produto obtido como resultado de nova amplificação da sequência nucleotídica sintética otimizada que codifica C, visando a eliminação da região C-terminal (hidrofóbica) revelou a presença de um fragmento de DNA com tamanho de 321pb (tamanho esperado=321pb; Figura 3.2 C).



**Figura 3.2.** Amplificação por PCR das regiões codificantes de C e NS3hel de sequências nucleotídicas de genótipo selvagem (A) e versões sintéticas com códons otimizados (B) e correspondente à região codificante de C de sequência nucleotídica sintética otimizada com uma deleção na região 3' (codificando um segmento hidrofóbico) (C), após separação por eletroforese em gel de agarose a 1,5%. M-NZYDNA Ladder VI. Controlos negativos: A - pistas 3 e 4, B - pistas 1 e 3, e C - pista 1. A - 1 e 2 indicam os fragmentos de DNA correspondentes às regiões codificantes de C e NS3hel, respetivamente. B- 2 e 4 indicam os fragmentos de DNA correspondentes às regiões codificantes de C e NS3hel, respetivamente. C- 2 indica o fragmento de DNA correspondente à região codificante de C.

### 3.3. Clonagem molecular das regiões que codificam C e NS3hel de ZIKV

Os fragmentos de DNA purificados (ver secção 2.7), correspondendo às regiões que codificam C (ou a sua variante com extremo C-terminal deletado) e NS3hel de ZIKV, amplificados de sequências nucleotídicas sintéticas otimizadas (ver secção 2.5.3) ou de genomas de ZIKV de genótipo selvagem (ver secção 2.5.2) foram clonados no vetor pET-29a usando *E. coli NovaBlue* como hospedeira de clonagem.

Os vetores pET (coleção disponibilizada comercialmente pela Novagen) apresentam a origem de replicação de pBR322 e estão presentes entre 15-20 cópias por célula. Estes plasmídeos apresentam um dos sistemas de promotores (promotor T7) mais eficientes quando se visa a expressão abundante de proteínas recombinantes, podendo estas representar até 50% da proteína celular total (Rosano e Ceccarelli, 2014). Nos vetores pET, o gene de interesse encontra-se a jusante do promotor T7, sendo este promotor de transcrição reconhecido pela polimerase de RNA do fago T7 e não pela polimerase de RNA de *E. coli*. Ainda que a polimerase de RNA do fago T7 possa ser expressa a partir de outro plasmídeo compatível com pET, tipicamente esta polimerase é expressa a partir

de uma cópia cromossomal do gene que a codifica, a qual se encontra integrada numa variante não-indutível do fago lambda (DE3) sob o controlo do promotor *lacUV5*, podendo assim ser induzido pela adição de IPTG (Studier e Moffatt, 1986; Novy e Morris, 2001). A jusante do promotor *T7*, no vetor utilizado, a existência de uma sequência do operador *lac* (*lacO*) permite regular firmemente a expressão basal a partir do promotor *T7* (aqui designado *T7-lac*) pela ligação do repressor LacI a *lacO*. Neste contexto, o repressor regula a expressão proteica heteróloga basal não só bloqueando o acesso da polimerase ao promotor *T7-lac* como pela repressão da transcrição a partir do gene que codifica a própria polimerase de RNA do fago T7, a qual é efetuada pela polimerase de *E. coli*, atuando sobre o promotor *lacUV5* no cromossoma da estirpe de expressão (Dubendorff e Studier, 1991). Assim, o sistema pET/DE3 permite contornar problemas de toxicidade, regulando firmemente o nível de expressão basal através do mecanismo de controlo descrito e entre outros, discutidos adiante.

Dependendo do conjunto de endonucleases de restrição utilizadas na digestão do vetor pET escolhido é possível a produção de proteínas recombinantes adicionadas em N- ou C-terminal de diferentes parceiros de fusão, os quais podem (i) permitir a fácil deteção das proteínas de interesse, e/ou (ii) a sua posterior purificação e/ou (iii) alterar a sua solubilidade. A escolha do conjunto de enzimas de restrição, *NotI* e *NdeI*, permitiu a adição em C-terminal de uma cauda de His<sub>6</sub>. O posicionamento de um parceiro de fusão em N-terminal e/ou C-terminal depende principalmente das características das proteínas de interesse.

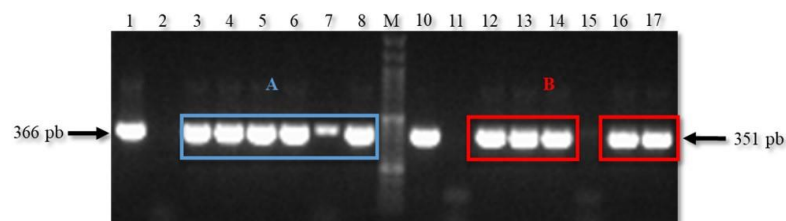
Neste estudo, o facto de as proteínas C e NS3hel apresentarem pequena massa molecular foi levado em conta face à decisão de colocação de His<sub>6</sub> em N- ou C-terminal. Tendo em conta as características dos vetores de expressão disponíveis optou-se pela colocação de um parceiro de fusão de pequenas dimensões em C-terminal, de forma a causar o mínimo impacto sobre o tamanho final do produto a produzir diminuindo, tanto quanto possível, a possibilidade de interferência do parceiro de fusão na aquisição de uma estrutura tridimensional correta pelas proteínas a produzir (e, idealmente, a sua solubilidade), bem como a atividade biológica e imunogenicidade das mesmas.

A construção dos vetores de expressão contendo os insertos das regiões codificantes de C e NS3hel, nas suas versões sintética com codões otimizados (*pET-29a::C(ot)* e *pET-29a::NS3hel(ot)*, respetivamente) e de genótipo selvagem (*pET-29a::C(wt)* e *pET-*

29a::NS3hel(wt), respetivamente), bem como, da região codificante de C de sequência nucleotídica sintética otimizada com região C-terminal (hidrofóbica) eliminada (*pET-29a::C(ot-del)*), foi iniciada através da realização de clivagem enzimática com endonucleases de restrição. O vetor pET-29a bem como os fragmentos de DNA amplificados por PCR (ver secções 2.5.2 e 2.5.3) e purificados (ver secção 2.7) foram duplamente digeridos com as enzimas de restrição *NotI* e *NdeI*. Após purificação dos produtos resultantes da digestão enzimática (ver secção 2.8.1.1) foram preparadas as reações de ligação entre o vetor pET-29a e os fragmentos de DNA usados como insertos (ver secção 2.8.2). As misturas de ligação foram utilizadas na transformação de células competentes de *E. coli NovaBlue* (ver secção 2.8.4). Após seleção de colónias contendo moléculas de DNA plasmídico possivelmente recombinante foi efetuada a sua extração pelo método de lise alcalina (ver secção 2.8.5). A confirmação da natureza recombinante das moléculas de DNA plasmídico foi efetuada através da amplificação, por PCR, dos insertos de DNA nelas contidos (ver secção 2.8.6).

A análise dos produtos obtidos como resultado das amplificações, quer para as sequências nucleotídicas de genótipo selvagem de ZIKV (resultados não apresentados), quer para as versões sintéticas com codões otimizados (Figura 3.3) revelou a presença de fragmentos de DNA com tamanho de 366pb para C, e 351pb para NS3hel. A análise dos produtos obtidos como resultado de nova amplificação da sequência nucleotídica sintética otimizada que codifica C, visando a eliminação da região C-terminal (hidrofóbica) revelou a presença de fragmentos de DNA com tamanho de 321pb (resultados não apresentados). De seguida, as moléculas de DNA recombinante purificadas foram sequenciadas parcialmente (região próxima do local de ligação do inserto ao vetor) e em ambos os sentidos, utilizando *primers* universais, complementares à sequência do vetor (ver secção 2.8.7). O processo de sequenciação permitiu constatar a inserção correta dos insertos de DNA no vetor de expressão, pela manutenção do local de ligação do ribossoma (rbs, do inglês *ribosome-binding site*) e da ORF que permite a expressão das proteínas C e NS3hel com as caudas de His<sub>6</sub> em C-terminal (resultados não apresentados). Apesar do seu carácter recombinante, ao longo do texto estas proteínas continuarão a ser referidas como C e NS3hel.





**Figura 3.3.** Rastreo por PCR de derivados de pET-29a contendo sequências de ZIKV, após separação por eletroforese em gel de agarose a 1,5%. M - *NZYDNA Ladder VI*. As caixas retangulares indicam os clones bacterianos onde foi detetada a presença de plasmídeos recombinantes. (A) 1 e (B) 10 – controles positivos; (A) 2 e (B) 11 – controles negativos; (A) e (B) - amplificação, por PCR, dos insertos correspondentes às regiões codificantes de C (3 a 8) e NS3hel (12 a 17) nas suas versões sintéticas com codões otimizados.

### 3.4. Expressão de proteínas recombinantes de ZIKV

Tal como referido acima, um dos fatores que, *a priori*, mais condicionaria a produção de C e NS3hel em quantidade seria o desajuste na leitura dos codões que as codificam naturalmente pela maquinaria de tradução de *E. coli*. Uma das abordagens para corrigir este problema passou pela otimização artificial das sequências das regiões codificantes dos polipéptidos C e NS3hel. Contudo, e uma vez que a preferência de utilização da população de codões reflete, aproximadamente, a abundância da população de tRNAs na célula, tal facto permite uma abordagem alternativa a este problema. Esta consiste na otimização do sistema de tradução pelo aumento da disponibilidade de tRNAs de baixa abundância no sistema, através da utilização de hospedeiros de expressão com cópias adicionais dos genes que codificam tRNAs que reconhecem codões raramente utilizados pelos mesmos, em plasmídeos (Gustafsson *et al.*, 2004; Sørensen e Mortensen, 2005b; Rosano e Ceccarelli, 2014). São destes exemplos as estirpes *Rosetta* (Novagen) ou *BL21-CodonPlus* (Agilent Technologies).

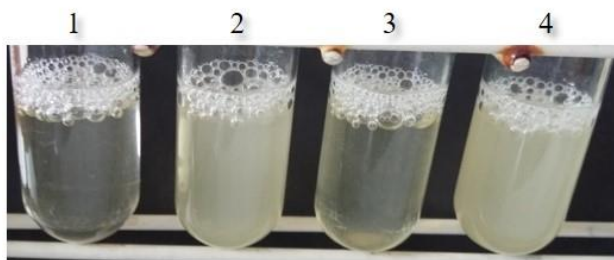
Neste estudo foram abordadas ambas as estratégias para contornar o desajuste na utilização de codões, tendo sido realizada a análise, em paralelo, da expressão proteica de sequências virais de genótipo selvagem e versões sintéticas com sequências otimizadas em diferentes estirpes de *E. coli*, *BL21(DE3)pLysS* e *Rosetta(DE3)pLysS*. As estirpes de *E. coli*, *BL21(DE3)pLysS* e *Rosetta(DE3)pLysS* ambas apresentam uma cópia cromossomal de um gene que codifica uma polimerase de RNA do fago T7 integrada no genoma de uma variante não indutível do fago lambda (DE3) e sob o controlo do promotor *lacUV5*, podendo assim ser induzido pela adição de IPTG (Studier e Moffatt, 1986; Novy e Morris, 2001). As duas estirpes apresentam, também, deficiência nas

proteases Lon e OmpT, o que diminui o risco de degradação das proteínas recombinantes nelas produzidas (Grodberg e Dunn, 1988). As duas estirpes contêm um plasmídeo (pLysS) que codifica um gene que confere resistência ao cloranfenicol e um gene que codifica a lisozima do fago T7, a qual é inibidora natural da polimerase de RNA do fago T7 (Moffatt e Studier, 1987; Studier, 1991). A enzima liga-se à polimerase impedindo a transcrição a partir do promotor *T7*, permitindo uma regulação firme do nível de expressão proteica basal. A estirpe *Rosetta(DE3)pLysS*, em especial, contém genes que codificam tRNAs que reconhecem códons raramente utilizados por *E. coli*. Os códons são AUA (Ile), AGG (Arg), AGA (Arg), CUA (Leu), CCC (Pro) e GGA (Gly). Os genes estão presentes no plasmídeo que transporta o gene que codifica a lisozima do fago T7, sendo uma variante de pLysS e designado pLysSRARE.

#### 3.4.1. Testes de expressão proteica e de otimização de produção de proteínas recombinantes

Os vetores de expressão contendo os insertos das regiões codificantes de C e NS3hel, nas suas versões sintética com códons otimizados (*pET-29a::C(ot)* e *pET-29a::NS3hel(ot)*, respetivamente) e de genótipo selvagem (*pET-29a::C(wt)* e *pET-29a::NS3hel(wt)*, respetivamente) foram utilizados na transformação de células competentes de *E. coli BL21(DE3)pLysS* e *Rosetta(DE3)pLysS* (ver secção 2.9). De seguida, foi efetuado o primeiro teste de expressão proteica, onde colónias bacterianas transformantes foram inoculadas em meio líquido e as suspensões bacterianas delas preparadas foram incubadas até uma D.O.<sub>600nm</sub>, de  $\approx 0,8$  e tratadas com IPTG (1mM final) durante 3h, a 37°C com agitação (200-220rpm) para induzir a expressão de proteínas heterólogas a partir dos derivados de pET-29a. Terminado o processo de indução foram colhidas alíquotas das culturas bacterianas transformadas com *pET-29a::NS3hel(ot)*, *pET-29a::NS3hel(wt)* e *pET-29a::C(wt)*, as quais foram ressuspensas em tampão de SDS-PAGE, e posteriormente analisadas através da separação das proteínas solúveis totais em géis de tricina-SDS-PAGE (ver secções 2.9.1 e 2.9.5). Contudo, quando induzida a expressão proteica heteróloga a partir da cultura bacteriana transportando *pET-29a::C(ot)* esta foi rapidamente acompanhada de lise da mesma, independentemente do contexto celular (*BL21* ou *Rosetta*; Figura 3.4). Esta lise deverá ter resultado, muito provavelmente, da destabilização da membrana celular por inserção da proteína C (devido

à presença de um extremo C-terminal com um domínio transmembranar) combinada com a ação da lisozima (codificada por pLysSRARE) libertada para o periplasma, promovendo assim a degradação do peptidoglicano e a consequente lise das células. Este efeito ocorreu apenas na cultura induzida, pelo que à partida, o nível de expressão proteica basal encontrava-se firmemente regulado pela ação dos mecanismos descritos acima (ação combinada do repressor LacI e da lisozima do fago T7).



**Figura 3.4.** Observação do efeito de toxicidade sobre culturas de *E. coli BL21(DE3)pLysS* e *Rosetta(DE3)pLysS* transformadas com *pET-29a::C(ot)*, quando induzida a expressão da proteína C, a 37°C durante 3h. 1 e 2 – culturas induzida (I) e não induzida (N.I) de *E. coli BL21(DE3)pLysS*, respetivamente. 3 e 4 - culturas I e N.I de *E. coli Rosetta(DE3)pLysS*, respetivamente.

A primeira abordagem para contornar este efeito foi implementada através da redução da taxa de síntese proteica, pela diminuição da temperatura de incubação (16°C e 25°C) e da concentração de IPTG utilizado (0,5mM final). Nas culturas bacterianas de *E. coli BL21(DE3)pLysS* e *Rosetta(DE3)pLysS* transformadas com *pET-29a::C(ot)*, a expressão proteica foi analisada promovendo a indução pretendida a 16°C ou 25°C com 0,5mM, durante a noite com agitação diminuída (170-180rpm) e a 25°C com 0,5mM, durante 4h e agitação regular (aproximadamente 220rpm). Em simultâneo foi realizada a análise da expressão proteica de culturas bacterianas transformadas com *pET-29a::C(wt)*, a 25°C com 0,5mM, durante a noite e agitação diminuída (170-180rpm). Findo o processo de indução da expressão proteica pretendida foram recolhidas alíquotas, as quais foram ressuspendidas em tampão de amostra e posteriormente analisadas por tricina-SDS-PAGE (ver secções 2.9.2 e 2.9.5). Inicialmente os extratos de proteínas solúveis totais foram analisados por eletroforese usando o método modificado de tricina-SDS-PAGE, descrito por Haider *et al.*, 2010. Uma vez que as proteínas C e NS3hel apresentam baixa massa molecular, o método indicado permite de forma fácil e rápida a separação deste tipo de proteínas num gel de baixa percentagem de poliacrilamida com maior resolução

que a observada nos géis de SDS-PAGE convencional (usando tampão eletrólito contendo glicina).

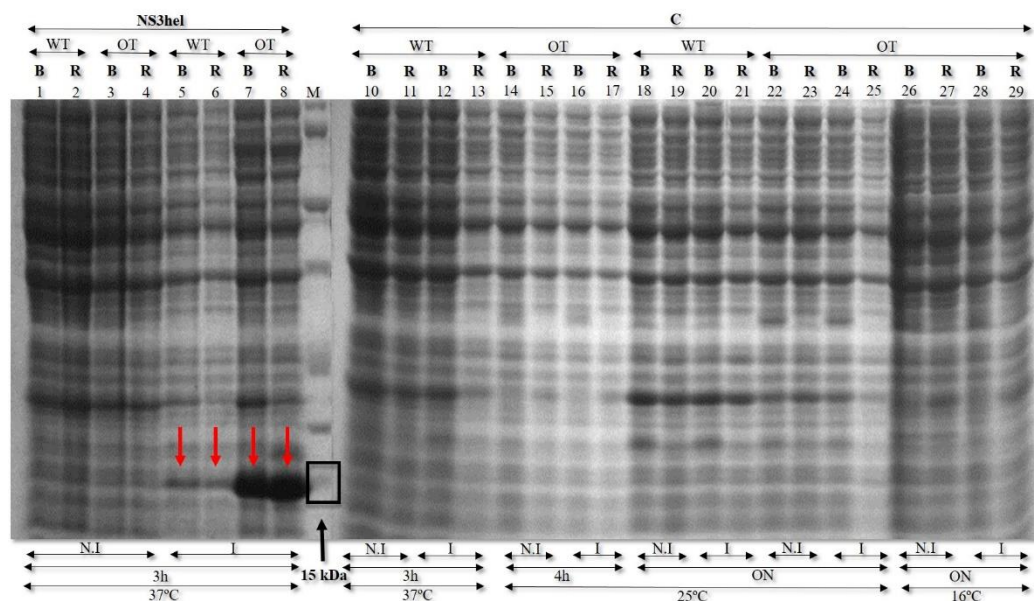
A análise eletroforética em géis de tricina-SDS-PAGE das condições testadas permitiu constatar, em primeiro lugar, que em todos os extratos proteicos totais derivados de culturas não induzidas não ocorrera produção das proteínas de interesse. Estes resultados possibilitam a confirmação do controlo do nível de expressão proteica basal pelos sistemas de expressão utilizados. A análise por tricina-SDS-PAGE do conjunto total de condições testadas, para a proteína NS3hel, revelou a presença de uma banda com  $\approx 15\text{kDa}$ , tal como teoricamente esperado. Os extratos proteicos totais derivados de culturas induzidas de *E. coli BL21(DE3)pLysS* e *Rosetta(DE3)pLysS* transformadas com *pET-29a::NS3hel(wt)* revelaram níveis de expressão proteica equivalentes (Figura 3.5). Ainda que visíveis em gel, a quantidade desta proteína nos extratos analisados era relativamente baixa. Assim, o aumento da disponibilidade de tRNAs de baixa abundância na estirpe *Rosetta(DE3)pLysS* não foi suficiente para contornar o desajuste na utilização de codões, não permitindo a otimização da expressão proteica. Tal facto não é, absolutamente surpreendente, na medida em que a estirpe em causa não dispõe de todos os tRNAs que reconhecem codões raramente utilizados por *E. coli*, nomeadamente, CGA e CGG, de entre os quais podem ser encontrados 2 codões CGG na sequência nucleotídica que codifica NS3hel.

Por oposição ao que se descreveu acima, para os extratos proteicos totais derivados de culturas de *E. coli BL21(DE3)pLysS* e *Rosetta(DE3)pLysS* transformadas com *pET-29a::NS3hel(ot)* [e, por conseguinte, expressando NS3hel(ot)], a indução da expressão proteica a partir de pET-29a revelou um aumento significativo nos níveis de expressão proteica heteróloga em relação às estirpes transformadas com *pET-29a::NS3hel(wt)* [e, por conseguinte, expressando NS3hel(wt)]. Entre as estirpes *BL21(DE3)pLysS* e *Rosetta(DE3)pLysS* não foram observadas diferenças significativas nos níveis de expressão proteica (Figura 3.5).

Assim neste estudo em particular, e no que à expressão de NS3hel diz respeito, a otimização artificial de sequências nucleotídicas permitiu corrigir o problema de desajuste na utilização de codões por *E. coli*, permitindo a produção da proteína pretendida com elevada eficiência. Ao longo do resto do trabalho optou-se por utilizar a estirpe *Rosetta(DE3)pLysS* transformada com *pET-29a::NS3hel(ot)*. A escolha de *E. coli*

*Rosetta(DE3)pLysS* parece algo redundante, já que, o aporte de tRNAs adicionais nada trás de novo a um sistema onde a expressão de NS3hel é feita a partir de uma sequência otimizada artificialmente. Contudo, a decisão de a utilizar teve em conta apenas o fator *logístico*, ou seja, *E. coli Rosetta(DE3)pLysS* atingiu mais rapidamente a fase de crescimento na qual a indução da expressão proteica foi efetuada.

Os extratos proteicos totais derivados de culturas induzidas de *E. coli BL21(DE3)pLysS* e *Rosetta(DE3)pLysS* expressando C(wt) não revelaram qualquer expressão proteica considerável, o qual, *a priori*, poderia ser expectável no caso da estirpe *BL21(DE3)pLysS* (Figura 3.5). Para *E. coli Rosetta(DE3)pLysS* era de esperar uma diferença significativa no nível de expressão proteica, no entanto, mais uma vez, a indisponibilidade de todos os tRNAs que reconhecem codões raramente utilizados por *E. coli* poderá ter comprometido a produção pretendida (1 codão CGG na sequência nucleotídica que codifica C). Outra hipótese colocada em discussão prende-se com a potencial toxicidade exercida pela proteína C sobre *E. coli*, sendo possível que as poucas células que possam estar a expressar C acabem eventualmente por lisar, e o extrato proteico total observado pode corresponder a células onde a produção não ocorreu. O primeiro conjunto de condições utilizado para contornar a toxicidade exercida pela proteína C não comprometeu por completo a viabilidade celular de culturas bacterianas transformadas com *pET-29a::C(ot)*, contudo, quando observados os extratos proteicos totais não se verificou qualquer expressão proteica considerável, pelo que também se colocou a hipótese de ocorrência de comprometimento da fisiologia nas células produtoras de proteína C (Figura 3.5). Também se colocou em consideração que as condições utilizadas (16°C ou 25°C com 0,5mM IPTG, durante a noite com agitação diminuída e a 25°C com 0,5mM IPTG, durante 4h e agitação regular) não tenham sido as suficientes para induzir uma produção proteica considerável. Os resultados obtidos levaram-nos a considerar a região C-terminal (hidrofóbica) como possivelmente estando na origem de alguns dos resultados obtidos, e a qual se determinou responsável pela toxicidade imposta sobre os hospedeiros de expressão.

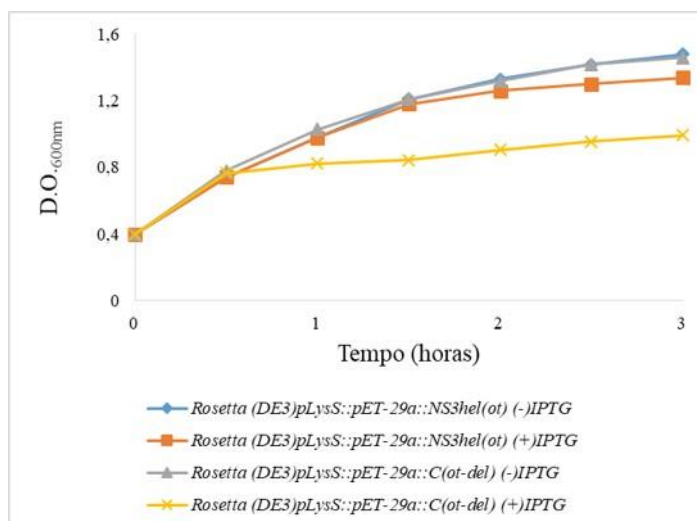


**Figura 3.5.** Análise de extratos de proteínas solúveis totais de culturas N.I e I de *E. coli BL21(DE3)pLysS* (B) e *Rosetta(DE3)pLysS* (R), expressando NS3hel(wt) (pistas 1, 2, 5 e 6), NS3hel(ot) (pistas 3, 4, 7 e 8), C(wt) (pistas 10-13; 18-21) e C(ot) (pistas 14-17; 22-29), em géis de tricina-SDS-PAGE (12%). Os tempos de indução e as temperaturas a que estas ocorreram estão indicados na figura. M-marcador de proteínas com massas moleculares conhecidas (*Precision Plus Protein™ Dual Color Standards*).

Após a construção do vetor de expressão contendo o inserto da região codificante de C de sequência nucleotídica sintética otimizada com região C-terminal (hidrofóbica) eliminada (*pET-29a::C(ot-del)*), procedeu-se à transformação de células competentes de *E. coli BL21(DE3)pLysS* e *Rosetta(DE3)pLysS* (ver secção 2.9). De seguida a expressão proteica heteróloga foi induzida através do tratamento das culturas com IPTG. Os extratos de proteínas solúveis totais obtidos foram analisados por eletroforese em géis de tricina-SDS-PAGE, de forma a avaliar a abordagem seguida para contorno do problema de toxicidade proteica. Em conjunto com esta análise foi ainda avaliado até que ponto a heterogeneidade da população bacteriana poderia comprometer o processo de expressão heteróloga. Para tal teve-se em conta a escolha de diferentes colónias bacterianas isoladas, as quais foram induzidas com IPTG (1mM final) durante 3h, a 37°C com agitação (200-220rpm). Desta vez, finda a indução da expressão pretendida não foram registadas alterações significativas na viabilidade do hospedeiro de expressão, tanto quanto estas podem ser traduzidas em variações da densidade ótica (discutido adiante). Findo o processo de indução da expressão proteica pretendida foram recolhidas alíquotas, as quais foram ressuspendidas em tampão de amostra e posteriormente analisadas por tricina-SDS-PAGE (ver secções 2.9.2 e 2.9.5).

Após eliminação da região C-terminal (hidrofóbica) na proteína C, ao contrário do inicialmente registado, aquando da indução da expressão da proteína C de ZIKV, a partir de variantes alélicas com codões otimizados para *E. coli*, as quais se traduziram na rápida lise da cultura após a adição de IPTG, a indução da expressão das variantes de C deletadas no seu extremo C-terminal nativo não mais induziram a lise do hospedeiro de produção. Assim, o problema de toxicidade proteica encontrava-se, *a priori*, resolvido. No entanto, para verificar se a expressão desta proteína recombinante ainda poderia estar a interferir na proliferação e homeostasia de *E. coli Rosetta(DE3)pLysS*, resultando apenas na diminuição da taxa de crescimento e densidade celular e não na lise celular, foi realizado um ensaio de avaliação da toxicidade proteica através da monitorização do crescimento bacteriano em meio líquido. Em conjunto também foi realizado o mesmo ensaio para a proteína NS3hel.

Idealmente, a realização de um ensaio de monitorização do crescimento de culturas de *E. coli Rosetta(DE3)pLysS*, não transformada e transformada com pET-29a teria permitido avaliar o *stress* metabólico induzido sobre *E. coli*, pela manutenção e replicação do plasmídeo. Além que um ensaio de monitorização do crescimento de culturas, antes de induzir, de *E. coli Rosetta(DE3)pLysS* não transformada, transformada apenas com pET-29a ou com *pET-29a::NS3hel(ot)* ou *pET-29a::C(ot-del)*, também teria permitido avaliar se os genes codificantes de NS3hel e C impõem algum efeito nefasto sobre o crescimento de *E. coli*. Contudo foi apenas avaliado a toxicidade decorrente da indução da expressão das proteínas heterólogas desejadas. Para tal, culturas bacterianas de *E. coli Rosetta(DE3)pLysS* expressando NS3hel(ot) ou C(ot-del) foram adicionadas de IPTG (1mM final) durante 3h, a 37°C com agitação (200-220rpm) de forma a induzir a expressão proteica heteróloga. No entanto foi adicionado IPTG quando atingida uma D.O.<sub>600nm</sub>, de  $\approx 0,4$ , em vez de  $\approx 0,8$ , de forma a impedir a saturação precoce das leituras de D.O. no espectrofotómetro utilizado. No momento do começo da indução foi iniciado o registo da D.O.<sub>600nm</sub> das diferentes culturas de 30 em 30min. Os valores de D.O.<sub>600nm</sub> obtidos foram utilizados na construção de curvas de crescimento para cada cultura (Figura 3.6).



**Figura 3.6.** Curvas de crescimento de culturas controlo (ausência de indução da expressão proteica heteróloga, indicadas por (-)IPTG ou de culturas em que a esta foi induzida pela adição de IPTG [indicadas por (+)IPTG]. A indução da expressão proteica pretendida foi iniciada a D.O.<sub>600nm</sub>, de ≈ 0,4, seguida de incubação a 37°C durante 3h.

A análise da Figura 3.6 permitiu verificar que as culturas não induzida e induzida de *E. coli* Rosetta(DE3)pLysS, expressando NS3hel(ot) apresentaram uma taxa de crescimento similar, pelo que, a indução da expressão proteica heteróloga não pareceu ter qualquer impacto no crescimento bacteriano.

Para a proteína C, a eliminação da região C-terminal não resolveu por completo o problema de toxicidade proteica, pois foi possível verificar, na Figura 3.6, uma variação decrescente na taxa de crescimento celular ao longo do tempo para a cultura induzida em relação à cultura que não foi adicionada de IPTG e, portanto, onde não foi induzida a expressão proteica heteróloga. Finda a indução, quando se procedeu à centrifugação para recolha das células também se verificou um menor rendimento na biomassa da cultura induzida expressando C(ot-del) comparando com a cultura não induzida expressando C(ot-del) e a induzida expressando NS3hel(ot), traduzido por uma diminuição evidente do tamanho do sedimento celular. Com estes resultados pode-se concluir que a proteína C exerce, quando comparada com NS3hel, maior *stress* fisiológico sobre o hospedeiro de expressão, levando a uma desaceleração no crescimento bacteriano. A estirpe Rosetta(DE3)pLysS não permite um nível de expressão ajustável e a quantidade de proteína produzida por célula é variável, pelo que possivelmente, a morte de subpopulações produtoras de elevadas quantidades de proteína C explica o menor rendimento total de biomassa (Rosano e Ceccarelli, 2014).

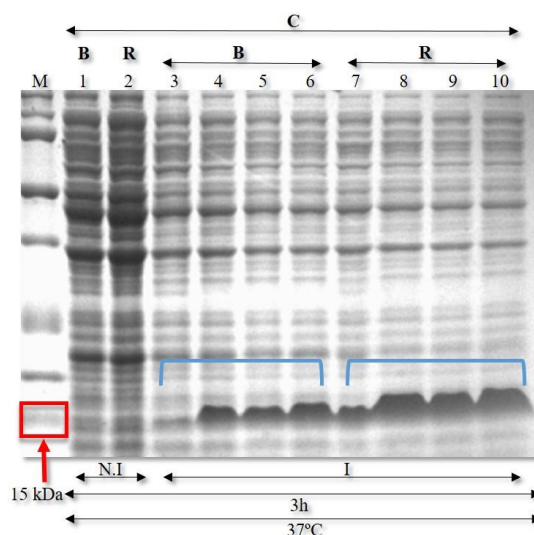


Estes resultados possibilitaram, mais uma vez, confirmar o controlo do nível de expressão proteica basal pelos sistemas de expressão utilizados, pois as culturas não induzidas expressando NS3hel(ot) ou C(ot-del) apresentaram curvas de crescimento similares, observando-se o efeito da expressão proteica apenas sobre a cultura induzida produtora da proteína C.

A análise eletroforética dos extratos de proteínas solúveis totais obtidos após induzida a produção da variante recombinante da proteína C de ZIKV com cauda de hexa-histidina revelou a presença de uma banda com  $\approx 13\text{kDa}$ , a qual corresponderia ao produto teoricamente esperado, e foi claramente observada nos extratos proteicos totais preparados após indução da expressão proteica a partir de culturas de *E. coli* BL21(DE3)pLysS e de *E. coli* Rosetta(DE3)pLysS expressando C(ot-del) (Figura 3.7.). Assim sendo, a produção da proteína C foi conseguida usando duas abordagens combinadas. A primeira delas teve por objetivo permitir a otimização da expressão proteica por via da alteração da sua sequência codificante (alteração de codões nativos por codões sinónimos, prontamente reconhecidos por *E. coli*). A segunda traduziu-se na deleção de parte da sequência nucleotídica que codifica o extremo C-terminal (hidrofóbico) da proteína C, o qual não se encontra na proteína viral madura, e cuja presença conduz à lise da cultura transportando variantes de pLysS tratadas com IPTG minutos após a adição deste indutor.

O parâmetro, heterogeneidade da população bacteriana (Figura 3.7.) permitiu avaliar diferenças no rendimento de expressão proteica entre as estirpes BL21(DE3)pLysS e Rosetta(DE3)pLysS e entre clones bacterianos das mesmas. Primeiramente, a análise dos perfis eletroforéticos obtidos revelou uma intensidade semelhante entre bandas proteicas equivalentes nos extratos proteicos totais obtidos (controlo). Contudo, quando observada a banda proteica correspondente à proteína C verificou-se diferenças de intensidade não só entre estirpes, mas também entre clones bacterianos, apontando uma aparente disparidade no rendimento de produção entre os mesmos. De facto, perante condições semelhantes, as células bacterianas podem encontrar-se em estados fisiológicos diferentes, uma vez que a sua fisiologia é perturbada aquando da indução da expressão proteica heteróloga, podendo cada célula responder de maneira diferente a um novo *stress* (Chae *et al.*, 2017). Por outro lado, a natureza das estirpes utilizadas não permite ajustar o nível de expressão proteica através de uma regulação "fina" na quantidade do indutor

utilizado, uma vez que o IPTG pode entrar nas células por transporte ativo, através da Lac-permease. Aliás a expressão de permeases em *E. coli* é heterogênea e o número de permeases ativas em cada célula é altamente variável, levando a que a concentração de IPTG dentro da célula não seja equivalente à presente no meio de cultura (Fernandez-Castane *et al.*, 2012; Rosano e Ceccarelli, 2014). Assim, cada estirpe/clon/célula apresenta o seu próprio rendimento de produção.



**Figura 3.7.** Análise, após separação por tricina-SDS-PAGE a 12%, de extratos proteicos totais derivados de culturas N.I e I de *E. coli* *BL21(DE3)pLysS* (B: 1, 3-6) e *Rosetta(DE3)pLysS* (R: 2, 7-10), expressando C(ot-del), a 37°C durante 3h, para avaliação do parâmetro, heterogeneidade da população bacteriana. M - marcador de proteínas com massas moleculares conhecidas (*Precision Plus Protein™ Dual Color Standards*).

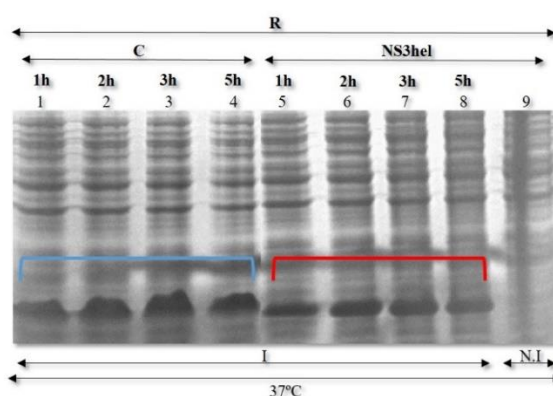
Após estes últimos resultados optou-se por dar continuidade ao estudo em questão com culturas de *E. coli* *Rosetta(DE3)pLysS* expressando C(ot-del). À semelhança do que já acima foi descrito para NS3hel, as estirpes derivadas de *Rosetta(DE3)pLysS* apresentaram consistentemente maiores níveis de expressão proteica, quando comparadas com as derivadas de estirpe *BL21(DE3)pLysS* (Figura 3.7), para além de atingirem, mais rapidamente, as densidades óticas às quais foram conduzidas as induções da expressão proteica pretendida.

Uma vez efetuada a escolha do hospedeiro para a expressão proteica que mais maximizava a produção das proteínas desejadas, para a otimização de produção das mesmas foi também efetuado um ensaio de cinética de produção, no qual, a expressão proteica heteróloga foi levada a cabo, a partir de culturas bacterianas de *E. coli*

*Rosetta(DE3)pLysS* expressando NS3hel(ot) ou C(ot-del), e pela adição de IPTG (1mM final), tendo estas, de seguida, sido mantidas a 37°C com agitação (200-220rpm). A tempos definidos após a adição de IPTG (1h, 2h, 3h e 5h) procedeu-se à recolha de alíquotas das suspensões bacterianas, as quais foram ressuspensas em tampão de amostra e posteriormente analisadas por eletroforese em géis de tricina-SDS-PAGE (ver secções 2.9.2 e 2.9.5).

A análise dos extratos proteicos totais obtidos revelou a saturação do processo de expressão proteica 2h a 3h após a adição de IPTG. De facto, quando observada a intensidade das bandas proteicas correspondentes às proteínas NS3hel ( $\approx 15$ kDa) e C ( $\approx 13$ kDa), nos diferentes momentos, verificou-se um aumento de intensidade entre a 1h e 2h. Contudo, entre as 2h e 3h e as 3h e 5h, as bandas proteicas em questão apresentaram uma intensidade semelhante (Figura 3.8).

Assim após avaliação de todos os parâmetros, optou-se nas fases seguintes, induzir a expressão das proteínas NS3hel e C com IPTG (1mM final) e incubação até uma D.O.<sub>600nm</sub>, de  $\approx 0,8$ , durante 3h, a 37°C com agitação (200-220rpm) (condição padrão).



**Figura 3.8.** Análise, após separação por tricina-SDS-PAGE a 12%, de extratos proteicos totais derivados de culturas N.I (9) e I de *E. coli Rosetta(DE3)pLysS* (R) expressando NS3hel(ot) (5-9) ou C(ot-del) (1-4), a 37°C durante 1h, 2h, 3h e 5h, para avaliação do parâmetro, ensaio de cinética de produção.

### 3.4.2. Testes de otimização de solubilidade proteica

A expressão de proteínas recombinantes no microambiente de *E. coli* difere do local onde originalmente a expressão natural de proteínas codificadas por ZIKV ocorre, *i.e.*, células de mosquitos e vertebrados suscetíveis a este vírus. Por isso mesmo, parâmetros como o pH, a osmolaridade, o potencial redox, a presença/ausência de cofatores e os mecanismos que determinam a aquisição de estrutura tridimensional de uma qualquer

proteína também são diferentes neste ambiente. Alterações nestes parâmetros, bem como a natureza intrínseca das proteínas recombinantes e a taxa de expressão a que são produzidas, podem levar a agregação proteica, resultando na formação dos chamados corpos de inclusão (Hartley e Kane, 1988; Carrió e Villaverde, 2001; Rosano e Ceccarelli, 2014).

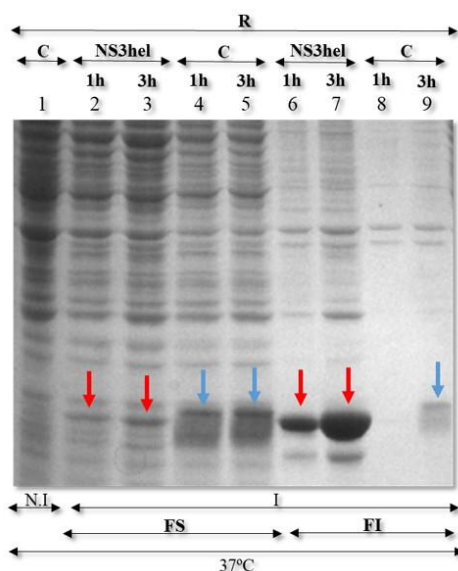
Tendo em conta que os fatores acima mencionados podem conduzir à acumulação de proteína heterologamente expressa em *E. coli* sob uma forma insolúvel foi realizado um ensaio de verificação da solubilidade das proteínas pretendidas, no qual culturas bacterianas de *E. coli Rosetta(DE3)pLysS* expressaram NS3hel(ot) ou C(ot-del) a 37°C com agitação (200-220rpm) durante 1h ou 3h, após tratamento com IPTG (1mM final). Os extratos proteicos obtidos, correspondendo às frações solúvel (sobrenadante) e insolúvel (sedimento), obtidas após lise celular e fracionamento por centrifugação, foram analisados por SDS-PAGE (ver secções 2.9.4 e 2.9.5), permitindo revelar, em ambas as frações, a presença de bandas com  $\approx 15\text{kDa}$  e  $\approx 13\text{kDa}$ , correspondendo a NS3hel e C-del, respetivamente (Figura 3.9).

Apesar da análise até aqui efetuada se ter baseado na utilização de géis de tricina-SDS-PAGE, entrando na fase de otimização da solubilidade das proteínas recombinantes foi sempre realizado o método convencional SDS-PAGE (Laemmli, 1970). Esta decisão foi tomada tendo em conta que não foi observado um decréscimo significativo na resolução da separação das proteínas NS3hel e C quando utilizando o método convencional.

É de notar que a migração eletroforética da proteína C foi ligeiramente diferente relativamente ao esperado, uma vez que, a banda correspondente à mesma, independentemente da técnica cromatográfica utilizada, encontra-se deslocada, acima da proteína NS3hel, tendo esta última uma massa molecular superior. O método de separação electroforética em géis de poliacrilamida (SDS-PAGE) permite a separação de proteínas com base na sua massa molecular num gel de poliacrilamida. O detergente, SDS, em combinação com um agente redutor e pela ação do calor interage com as proteínas, desnaturando-as e conferindo-lhes uma carga global negativa. Nestas condições, teoricamente, a massa molecular seria o único fator a afetar a migração eletroforética. Contudo, o fenómeno de migração aberrante não é raro, podendo este caso ser explicado pela má interação do SDS com a proteína C, fazendo com que a migração no campo

elétrico seja mais lento. Outros fatores podem também ter contribuído para o fenômeno observado, tais como, a presença no tampão de uma elevada concentração de sal e/ou do detergente Triton X-100, que pode ter competido com o SDS pela interação com a proteína C.

A análise também permitiu verificar que a proteína NS3hel apresenta-se maioritariamente na fração insolúvel, enquanto a proteína C está maioritariamente na fração solúvel. Também se pode observar, que para a proteína NS3hel o aumento de tempo de indução, de 1h para 3h, resulta no aumento de produção de proteína sob a forma insolúvel. De facto, se bem que se tenham observado bandas de intensidade semelhante nas proteínas presentes nas frações solúveis preparadas a partir de culturas tratadas com IPTG durante 1h ou 3h, comparativamente foi observado um aumento significativo entre as intensidades das bandas correspondentes a NS3hel nas frações insolúveis respetivas (Figura 3.9.).



**Figura 3.9.** Análise, após separação por SDS-PAGE a 13%, de um extrato de proteínas solúveis totais de uma cultura controlo (ausência de indução da expressão proteica heteróloga; 1) e de extratos proteicos correspondendo às frações solúvel (FS) e insolúvel (FI), de culturas onde foi induzida a produção heteróloga de NS3hel (2, 3, 6 e 7) ou C (4, 5, 8 e 9), a 37°C durante 1h e 3h, para verificação da solubilidade proteica.

Tal como acima se menciona, a sobre-expressão de proteínas recombinantes em *E. coli* resulta, frequentemente, na produção de proteínas sob a forma insolúvel, sob a forma de corpos de inclusão (CI). A produção de proteínas heterólogas sob a forma de CI ocorre como resposta a uma acumulação de proteína geralmente com uma estrutura tridimensional incorreta, em consequência de uma velocidade de síntese proteica superior

à capacidade de aquisição de uma conformação (*folding*) correta, provavelmente decorrente de incapacidade de interação com os moduladores do processo de *folding*, tais como chaperones moleculares, que neste tipo de situação ficam rapidamente saturados. Quando as proteínas não se encontram na sua estrutura tridimensional correta apresentam, usualmente regiões hidrofóbicas expostas, as quais podem interagir com outras proteínas semelhantes, formando agregados proteicos. Este problema é muito comum no processo de expressão heteróloga, especialmente se conduzido a partir de promotores fortes, contudo não existe estabelecida uma estratégia universal que previna a agregação proteica (Baneyx e Mujacic, 2004; Sørensen e Mortensen, 2005a).

Tal como referido acima, este estudo visou alcançar uma produção de variantes recombinantes de uma proteína viral (C) e um domínio funcional (hel) de uma outra (NS3) sendo que, idealmente, estes devam apresentar-se nos extratos obtidos, o mais semelhantes possível aos produzidos pelas células permissivas infetadas. Para além dos ensaios funcionais (por exemplo testando a ligação de C ao RNA viral ou da atividade de helicase de NS3hel) o exigirem, é facilitada ainda a produção de anticorpos capazes de as detetar *in vivo*. De facto, aquando da produção de anticorpos, a apresentação antigénica deverá ser o mais semelhante possível ao processo de resposta natural. Se assim for, os soros policlonais produzidos encontrar-se-ão enriquecidos em anticorpos específicos contra epítomos/determinantes de superfície das proteínas nativas, garantido uma maior segurança de reconhecimento das proteínas virais no contexto dos estudos de estratégia replicativa em células permissivas (*i.e.*, num contexto em que as proteínas virais são detetadas na sua forma nativa). Em alternativa, a produção de soros policlonais preparados a partir de proteínas desnaturadas pode apresentar-se enriquecida naqueles que reconhecem apenas epítomos lineares, os quais podem, ou não, ser expostos à superfície destas moléculas quando expressas *in vivo*.

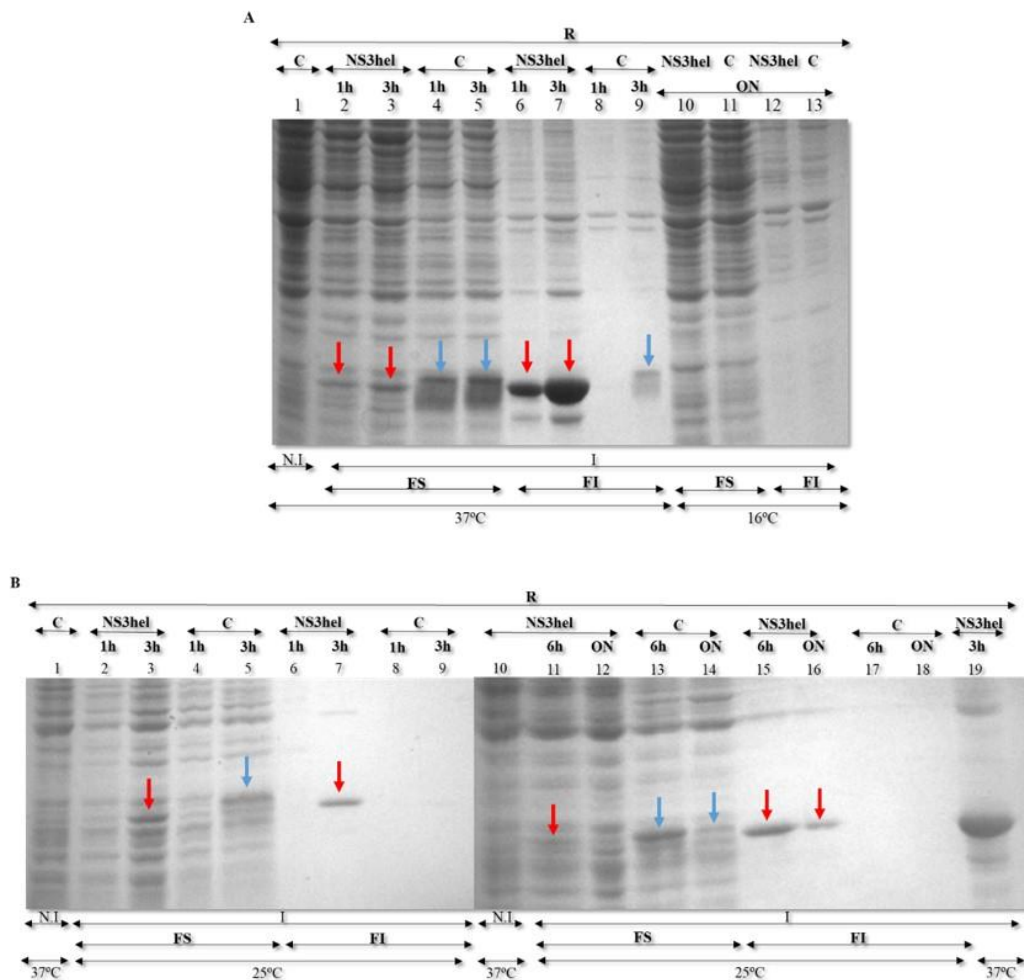
No presente estudo foram avaliados vários parâmetros de forma a melhorar a solubilidade das proteínas recombinantes, os quais incluíram: (i) a variação da temperatura de crescimento, (ii) a alteração do tempo de indução e (iii) da fase de crescimento na qual a indução da expressão proteica foi efetuada, (iv) o uso do aditivo D-sorbitol no meio de crescimento e (v) o mecanismo de lise celular utilizado.

Para limitar a agregação *in vivo*, primeiramente, a taxa de síntese proteica foi reduzida pela diminuição da temperatura de incubação (16°C e 25°C em vez de 37°C). Culturas

bacterianas de *E. coli* Rosetta(DE3)pLysS expressando NS3hel(ot) ou C(ot-del) foram tratadas com IPTG (1mM final) a 16°C ou 25°C com agitação diminuída (170-180rpm) durante a noite. Em simultâneo foi realizada a análise da expressão proteica heteróloga de culturas bacterianas a 25°C durante 1h, 3h e 6h (ver secção 2.9.4). Os extratos proteicos obtidos, correspondendo às frações solúvel (sobrenadante) e insolúvel (sedimento), foram analisados por SDS-PAGE (ver secção 2.9.5; Figura 3.10). Os extratos proteicos derivados de culturas induzidas a 16°C não revelaram qualquer expressão proteica considerável, correspondente às proteínas de interesse, em ambas as frações. Na cultura onde NS3hel foi expressa a 25°C durante 3h, foi possível verificar no seu extrato proteico, correspondendo à fração solúvel, uma produção semelhante ao extrato proteico derivado de uma cultura mantida a 37°C (condição padrão). Quando se efetuou a produção a 25°C durante 6h verificou-se um nível de expressão menor comparando com a indução na condição padrão. Quando se realizaram ensaios de expressão proteica à temperatura de 25°C, em que esta foi estimulada durante 1h e durante a noite, não foi registada qualquer expressão proteica heteróloga considerável sob a forma de proteína recombinante solúvel, verificando-se apenas produção heteróloga, sob a forma insolúvel, após uma noite de incubação. Nas culturas produtoras de proteína C, a 25°C durante 3h e 6h, também foi possível revelar nos seus extratos proteicos, correspondentes à fração solúvel, uma produção semelhante com a indução a 37°C e 3h, enquanto que a indução durante a noite obteve um nível de expressão proteica menor. Quando se induziu a produção da proteína C a 25°C durante 1h não se registou uma produção considerável desta, e independentemente da sua solubilidade (Figura 3.10.).

Teoricamente, as baixas temperaturas contribuem para uma diminuição da taxa de transcrição e tradução e levam a uma redução da força imposta pelas interações hidrofóbicas, permitindo que as proteínas adquiram uma estrutura tridimensional correta. No entanto, a mesma abordagem também leva a uma redução na produtividade decorrente de uma diminuição da eficiência/velocidade de todos os processos metabólicos, traduzidos, inclusive, em menor taxa de crescimento celular a estas temperaturas, o que resulta na diminuição da biomassa final (Baneyx e Mujacic, 2004). Estes efeitos foram observados no presente estudo, uma vez que, por um lado, não foi observada a produção das proteínas desejadas a 16°C e, por outro, a 25°C não houve qualquer aumento na expressão proteica final, sob a forma solúvel, em comparação com a indução na condição

padrão. Para NS3hel, quando efetuada a sua produção heteróloga a 25°C, apesar de se ter notado uma menor produção sob a forma insolúvel, esta não se refletiu no aumento de produção sob a forma solúvel. Aliás, para ambas as proteínas, observou-se uma redução no nível de produção à medida que se aumentou o tempo de indução (redução drástica entre uma indução a 3h e durante a noite). Para o explicar colocou-se como hipótese que a estabilidade das proteínas foi afetada pela longa incubação.



**Figura 3.10.** Análise, após separação por SDS-PAGE a 13%, **A)** de um extrato de proteínas solúveis totais de uma cultura controlo (1) e de extratos proteicos correspondendo às frações solúvel (FS) e insolúvel (FI), de culturas onde foi induzida a produção heteróloga de NS3hel (2, 3, 6, 7, 10 e 11) ou C (4, 5, 8, 9, 12 e 13). **B)** de extratos de proteínas solúveis totais de culturas controlo (1 e 10) e de uma cultura onde foi induzida a produção heteróloga de NS3hel (19), e de extratos proteicos correspondendo às frações solúvel e insolúvel, de culturas onde foi induzida a produção heteróloga de NS3hel (2, 3, 6, 7, 11, 12, 15 e 16) ou C (4, 5, 8, 9, 13, 14, 17 e 18). Os tempos de indução e as temperaturas a que estas ocorreram estão indicados na figura.

A expressão proteica heteróloga é induzida, por norma, a uma D.O.<sub>600nm</sub> oscilando entre 0,4 e 1, uma vez que neste intervalo, correspondendo ao meio da fase exponencial

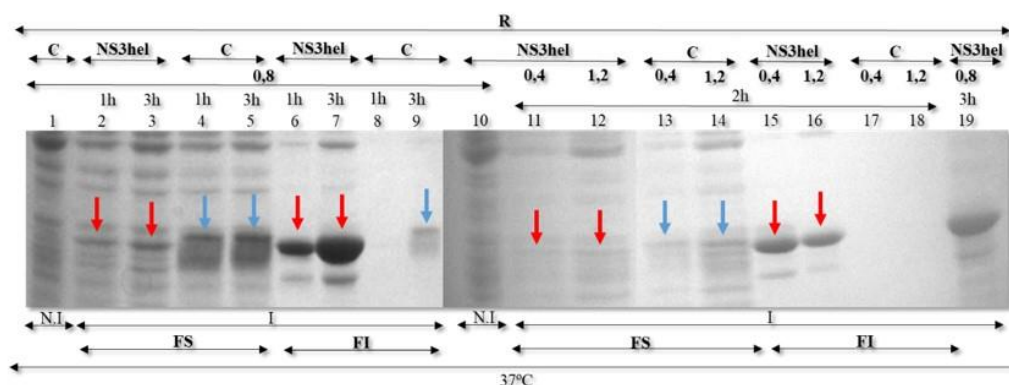


(do inglês *mid-log phase*), a taxa de crescimento e o processo de tradução encontram-se no seu potencial máximo, permitindo um elevado rendimento do processo de expressão proteica. No entanto, e frequentemente, o elevado nível de expressão proteica não se reflete na produção de proteínas com elevado grau de solubilidade. Por isso, numa nova tentativa de melhorar a solubilidade das proteínas recombinantes foi avaliado um novo parâmetro, neste caso a fase de crescimento na qual a indução da expressão proteica foi efetuada.

De uma forma geral, estes ensaios pretendem, uma vez mais, reduzir a taxa de síntese proteica. Para isso foram testadas duas induções da expressão proteica a novas D.O.<sub>600nm</sub> (0,4 e 1,2), correspondendo, ao início do meio da fase exponencial e fim da fase exponencial de crescimento, respetivamente. Culturas bacterianas de *E. coli Rosetta(DE3)pLysS*, expressando NS3hel(ot) ou C(ot-del) foram induzidas a D.O.<sub>600nm</sub>, de  $\approx 0,4$  e 1,2 com IPTG (1mM final), a 37°C durante 2h com agitação (200-220rpm). A análise dos extratos proteicos obtidos permitiu verificar que não ocorreu qualquer aumento de produção das proteínas C e NS3hel, sob a forma solúvel, comparando com a indução a D.O.<sub>600nm</sub>, de  $\approx 0,8$  com IPTG (1mM final), a 37°C durante 3h com agitação (200-220rpm) (condição padrão; Figura 3.11). Aliás, para ambas as proteínas, observou-se uma redução no nível de produção, em ambas as frações. De facto, para NS3hel, apesar de se ter notado uma menor produção sob a forma insolúvel, esta não se refletiu no aumento de produção sob a forma solúvel (Figura 3.11).

Numa indução a D.O.<sub>600nm</sub> de  $\approx 0,4$ , as células estão apenas no início do período onde a taxa de crescimento e o processo de tradução atingem o seu potencial máximo (o qual corresponde a uma D.O.<sub>600nm</sub> oscilando entre 0,6-0,8). Assim é permitida a redução da taxa de síntese proteica nesta D.O., o que poderia possibilitar a aquisição de uma conformação (*folding*) correta das proteínas de interesse e limitar a agregação proteica (San-Miguel *et al.*, 2013). Contudo, a indução da expressão proteica heteróloga leva a uma diminuição da taxa de crescimento, uma vez que a maquinaria celular está centrada na produção de proteína pretendida. Por se tratar de um período inicial de crescimento celular, o nível de biomassa total é baixa, o que leva no final a um menor rendimento proteico total, comparando com a indução na condição padrão (onde os níveis de biomassa total e do processo de tradução estão no seu ideal). A indução no fim da fase exponencial representa um período onde ocorre uma diminuição da taxa de crescimento,

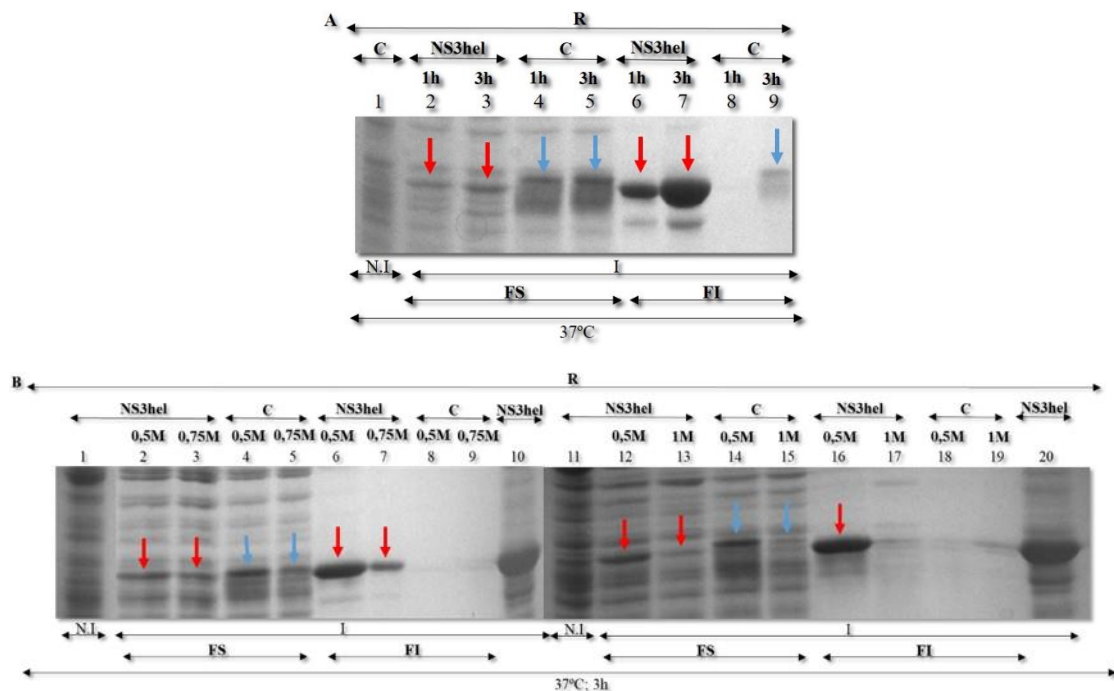
e por isso, a taxa de síntese proteica também é reduzida, o que poderia limitar a agregação proteica e formação de CI (Galloway *et al*, 2003). Assim, o problema associado a insolubilidade proteica não foi resolvido por alteração do parâmetro em causa e a diminuída produção heteróloga, nestes dois períodos, pode ter ocorrido, para uma D.O.<sub>600nm</sub>, de  $\approx 0,4$  devido a uma baixa biomassa total e a  $\approx 1,2$  devido a mudanças metabólicas associadas a este período que resultam num limitado crescimento bacteriano e aumento do número de células não viáveis.



**Figura 3.11.** Análise, após separação por SDS-PAGE a 13%, de extratos de proteínas solúveis totais de culturas controlo (1 e 10) e de uma cultura onde foi induzida a produção heteróloga de NS3hel (19), e de extratos proteicos correspondendo às frações solúvel (FS) e insolúvel (FI), de culturas onde foi induzida a produção heteróloga de NS3hel (2, 3, 6, 7, 11, 12, 15 e 16) ou C (4, 5, 8, 9, 13, 14, 17 e 18). A indução da expressão proteica pretendida foi iniciada a D.O.<sub>600nm</sub>, de  $\approx 0,4$ ,  $0,8$  ou  $1,2$  seguida de incubação a  $37^{\circ}\text{C}$  durante 1h, 2h ou 3h.

De seguida, a adição ao meio LB líquido do aditivo D-sorbitol foi avaliada como nova abordagem de melhoramento da solubilidade proteica. O sorbitol é um osmólito (designado de chaperone químico) utilizado, geralmente, como estabilizador de proteínas *in vitro*. Quer isto dizer que, na sua presença, a alteração da estrutura tridimensional de proteínas corretamente estruturadas é inibida, forçando estas proteínas a manterem um estado conformacional nativo (Yang e Zhang, 2013). Neste estudo o efeito na solubilidade das proteínas heterólogas produzidas, decorrente da sua síntese na presença de D-sorbitol, foi avaliado através da adição deste composto nas concentrações de  $0,5\text{M}$ ,  $0,75\text{M}$  e  $1\text{M}$ . Culturas bacterianas de *E. coli Rosetta(DE3)pLysS*, expressando NS3hel(ot) ou C(ot-del) foram inoculadas em meio LB líquido suplementado com D-sorbitol e, em seguida, a expressão proteica heteróloga induzida na condição padrão. A análise efetuada permitiu verificar que, para uma concentração de  $0,5\text{M}$ , ocorreu um aumento de produção das proteínas C e NS3hel sob a forma solúvel comparando com a indução sem suplementação

(Figura 3.12). Desta vez, para NS3hel, uma menor produção sob a forma insolúvel refletiu-se no aumento de produção sob a forma solúvel. De facto, o D-sorbitol afeta negativamente a taxa de crescimento de *E. coli*, por isso, a uma concentração de 0,5M, o efeito sobre o crescimento parece igualmente refletir-se na redução da taxa de síntese proteica e, consequentemente, que poderá adicionar-se ao seu efeito de chaperone, contribuindo para o correto *folding* proteico e aumento de solubilidade (Yang e Zhang, 2013). Para as concentrações de 0,75M e 1M observou-se um decréscimo sucessivo na produção heteróloga de ambas as proteínas na fração solúvel e para NS3hel também na fração insolúvel (Figura 3.12 B). Aliás, em meio de cultura contendo 1M de D-sorbitol foi diminuída a síntese proteica ao ponto que não foi registada qualquer produção considerável sob a forma insolúvel, e a pouca que foi detetada encontra-se na fração solúvel (Figura 3.12 B). A diminuição da produção das proteínas desejadas pode ser explicada pelo elevado *stress* osmótico imposto pelo D-sorbitol, que resultou num limitado crescimento bacteriano e baixa produção heteróloga.



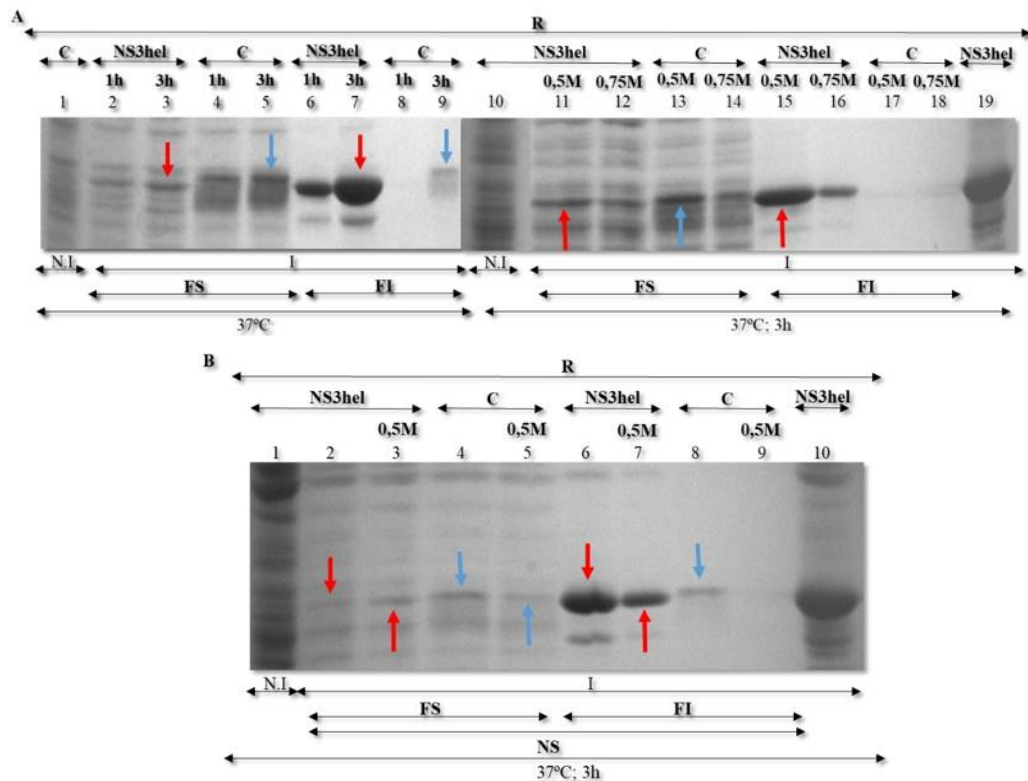
**Figura 3.12.** Análise, após separação por SDS-PAGE a 13% **A)** de um extrato de proteínas solúveis totais de uma cultura controlo (1) e de extratos proteicos correspondendo às frações solúvel (FS) e insolúvel (FI), de culturas onde foi induzida a produção heteróloga de NS3hel (2, 3, 6 e 7) ou C (4, 5, 8 e 9). **B)** de extratos de proteínas solúveis totais de culturas controlo (1 e 11) e de culturas onde foi induzida a produção heteróloga de NS3hel (10 e 20), e de extratos proteicos correspondendo às frações solúvel e insolúvel, de culturas onde foi induzida a produção heteróloga de NS3hel (2, 3, 6, 7, 12, 13, 16 e 17) ou C (4, 5, 8, 9, 14, 15, 18 e 19). As concentrações de D-sorbitol testadas, os tempos de indução e a temperatura a que as induções ocorreram estão indicados na figura.

Assim após avaliação de todos os parâmetros acima mencionados optou-se, nas fases seguintes deste trabalho, pela indução da expressão de C e NS3hel de ZIKV com 1mM de IPTG em meio LB líquido suplementado com 0,5M D-sorbitol e incubação até uma D.O.<sub>600nm</sub>, de  $\approx 0,8$ , durante 3h, a 37°C com agitação (200-220rpm), uma vez que, esta condição obteve melhor rendimento e qualidade de produção em relação a todas as condições testadas.

Finalmente, para avaliar o eventual impacto de lise celular na otimização de solubilidade proteica foram utilizados dois mecanismos de lise. Neste caso foram testados 5 ciclos de congelamento/descongelamento em azoto líquido/37°C e a sonicação. Ambos os mecanismos correspondem a métodos físicos de lise celular, que no caso do processo de sonicação origina ondas sonoras pulsadas e de alta frequência, enquanto que durante os ciclos repetidos de congelamento/descongelamento em azoto líquido/37°C a lise das células bacterianas decorre da expansão/contração de cristais de gelo formados durante este processo. Um dos problemas que advém da utilização da sonicação é a geração de calor localizado no interior das amostras, e de forma a evitar ao máximo este problema, estas são mantidas em banho de gelo durante todo o processo. No entanto, ainda assim este calor pode contribuir para a desnaturação e agregação proteica.

O conjunto de 5 ciclos de congelamento/descongelamento seguidos de sonicação foi utilizado em todos os ensaios até agora descritos. A utilização exclusiva de 5 ciclos de congelamento/descongelamento foi testada com a condição padrão de indução e com a condição que obteve maior grau de solubilidade proteica até ao momento (tal como descrito acima). Este último ensaio serviu para avaliar, apesar de saber-se *a priori* que a fração insolúvel estaria bastante enriquecida em CI e células não lisadas, se de alguma forma o processo de sonicação poderia estar a contribuir para a insolubilidade proteica observada. Tendo isto em conta, pretendia-se testar se ocorreria um aumento de quantidade da proteína NS3hel sob a forma solúvel em relação às condições onde o processo de sonicação foi empregue. Contudo, a análise permitiu verificar que, quando a lise celular foi promovida exclusivamente com ciclos de congelamento/descongelamento, ocorreu uma diminuição de intensidade das bandas correspondentes às proteínas C e NS3hel (quer na forma solúvel, quer insolúvel) quando comparando com o obtido aquando da promoção de lise celular pela ação combinada de um conjunto de 5 ciclos de congelamento/descongelamento seguidos de sonicação, em condições de indução idênticas

(Figura 3.13). A análise tem associada um fator de confundimento na medida em que os extratos proteicos de culturas cuja lise celular foi promovida pela combinação ciclos de congelação/descongelação-sonicação encontravam-se 10X concentrado (relativamente ao volume de cultura de partida) enquanto que os extratos proteicos produzidos a partir de culturas cuja lise foi promovida exclusivamente com ciclos de congelação/descongelação, estavam apenas concentrados 3X relativamente ao volume inicial. Apesar desta situação supõe-se que não ocorreu um aumento de produção sob a forma solúvel para NS3hel, e consequentemente a insolubilidade deverá advir da sua natureza proteica, ou seja, quando produzida no microambiente de *E. coli* tem tendência para se agregar sob a forma de CI. Assim exclui-se o processo de sonicação como fator crucial contribuinte para a insolubilidade proteica. Porém, poderia ter sido realizado um outro ensaio de forma a averiguar a contribuição do mecanismo ciclos repetidos de congelação/descongelação em azoto líquido/37°C para o fenómeno de insolubilidade proteica, em particular o passo de descongelação a 37°C, pelo facto da agregação proteica ser favorecida a estas temperaturas. Em futuros estudos, como alternativa, um outro método de lise celular pode ser testado, a prensa de *French*.



**Figura 3.13.** Análise, após separação por SDS-PAGE a 13% A) de extratos de proteínas solúveis totais de culturas controlo (1 e 10) e de uma cultura onde foi induzida a produção heteróloga de NS3hel (19) e de extratos proteicos correspondendo às frações solúvel (FS) e insolúvel (FI), de culturas onde foi induzida a

produção heteróloga de NS3hel (2, 3, 6, 7, 11, 12, 15 e 16) ou C (4, 5, 8, 9, 13, 14, 17 e 18). **B)** de extratos de proteínas solúveis totais derivados de uma cultura controle (1) e de uma onde foi induzida a produção heteróloga de NS3hel (10), e de extratos proteicos, obtidos sem recurso a sonicação (NS), correspondendo às frações solúvel e insolúvel, de culturas onde foi induzida a produção heteróloga de NS3hel (2, 3, 6 e 7) ou C (4, 5, 8 e 9). As concentrações de D-sorbitol testadas, os tempos de indução e a temperatura a que as induções ocorreram estão indicados na figura.

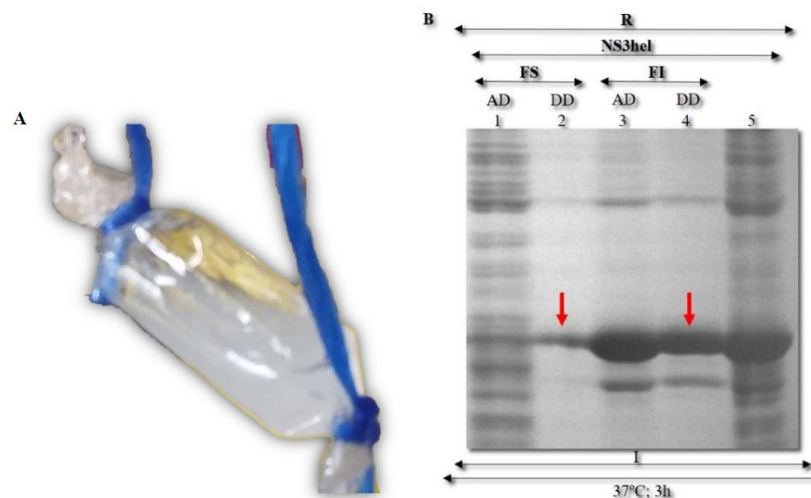
### **3.4.3. Recuperação da estrutura tridimensional de proteínas desnaturadas (*refolding* proteico)**

As estratégias utilizadas até ao momento tiveram como objetivo principal permitir a obtenção de um extrato proteico enriquecido nas proteínas de interesse, tanto quanto possível, na sua forma solúvel. No entanto, uma vez que os CI podem conter a proteína pura e intacta (ainda que desnaturada), pode optar-se por proceder à sua reestruturação tridimensional (*refolding*) *in vitro*, sendo que, nalguns casos pelo menos, é possível a obtenção de proteína na conformação nativa a partir de um extrato contendo essencialmente proteína agregada em CI. No decurso deste trabalho optou-se por abordar esta possibilidade usando a diálise, que permite o *refolding* proteico direto a partir de extratos de proteínas desnaturadas ou com uma conformação tridimensional incorreta, acumuladas e agregadas sob a forma de CI, as quais foram solubilizadas *in vitro* pela sua desnaturação na presença de ureia.

Para tal, NS3hel(ot) foi produzida, nas condições padrão, usando uma cultura bacteriana de *E. coli Rosetta(DE3)pLysS*. O *refolding* proteico direto, utilizando o processo de diálise, foi permitido pela remoção gradual da ureia, que através da sua ação caotrópica permitiu a solubilização das proteínas presentes nos extratos proteicos. Para tal, o extrato proteico obtido correspondendo à fração insolúvel foi solubilizado recorrendo a uma solução tampão com 8M de ureia e de seguida submetido a um conjunto de diálises sucessivas, com recurso a três tampões com concentrações decrescentes de ureia (4M e 2M), com o último tampão a não apresentar na sua composição qualquer ureia. A análise permitiu verificar a ocorrência de um aumento de intensidade da banda correspondente à fração solúvel da proteína NS3hel com diminuição na fração insolúvel (Figura 3.14 B). No entanto, esse aumento não foi considerado significativo, continuando a maioria da proteína NS3hel sob a forma insolúvel. Este resultado não foi considerado surpreendente, pois na maioria dos casos, a diminuição da concentração de desnaturante no decurso do processo de diálise, quando ocorre abaixo de um nível considerado insuficiente para garantir a desnaturação das proteínas presentes, pode traduzir-se na sua

precipitação, especialmente se a concentração da proteína for elevada, o que resulta no baixo rendimento de recuperação proteica.

Apesar do seu potencial, o *refolding* proteico é um processo que requer uma elevada otimização, o que por sua vez necessita de um enorme consumo de tempo, do qual não se dispunha. Portanto, em futuros estudos a otimização do protocolo utilizado será necessário, por exemplo, através da adição de aditivos químicos ao tampão de *refolding*, como, aminoácidos, polietilenoglicol e ciclodextrinas, atuando como estabilizadores de proteínas ou inibidores da agregação proteica (Yamaguchi e Miyazaki, 2014).



**Figura 3.14.** A) Observação de um extrato proteico total depois do processo de diálise e o qual é derivado de uma cultura onde foi induzida a produção heteróloga de NS3hel. O aspeto “leitoso” da suspensão proteica é indicativo da precipitação de proteínas. B) Análise, após separação por SDS-PAGE a 13%, de um extrato proteico total (5) e de extratos proteicos correspondendo às frações solúvel (FS) e insolúvel (FI), antes (AD) e depois (DD) do processo de diálise, os quais são derivados de culturas onde foi induzida a produção heteróloga de NS3hel, a 37°C durante 3h.

#### 3.4.4. Purificação de proteínas em condições desnaturantes e nativas

Até ao momento, a produção das proteínas de fusão pretendidas foi assumida como tendo ocorrido com sucesso pela observação do aumento da intensidade de uma espécie proteica nos extratos produzidos, a qual se apresentava com a massa esperada para os produtos pretendidos. No entanto, tal facto é apenas evidência presumível da identidade das proteínas de fusão em questão, pelo que a identidade das proteínas de fusão C e NS3hel com caudas de His<sub>6</sub> na extremidade C-terminal foi subsequentemente investigada. Para tal foram utilizados dois métodos que visaram a identificação da presença de uma

cauda de hexa-histidina nas proteínas produzidas. O primeiro destes métodos correspondeu à purificação das mesmas por cromatografia de afinidade.

O tipo de cromatografia de afinidade escolhido fez uso da utilização de uma coluna que continha uma matriz na qual estavam imobilizados íons níquel. Por sua vez, as moléculas de histidina, funcionando como parceiro de fusão nas proteínas recombinantes produzidas, apresentam uma elevada afinidade pelos íons níquel, pois os grupos dadores de elétrons no seu anel de imidazole formam rapidamente ligações de coordenação com os íons níquel, permitindo a purificação após passagem de um extrato proteico enriquecido em proteína de fusão pela coluna de cromatografia (Bornhorst e Falke, 2000). A técnica foi realizada em condições desnaturantes e em condições nativas (ver secções 2.9.7 e 2.9.8).

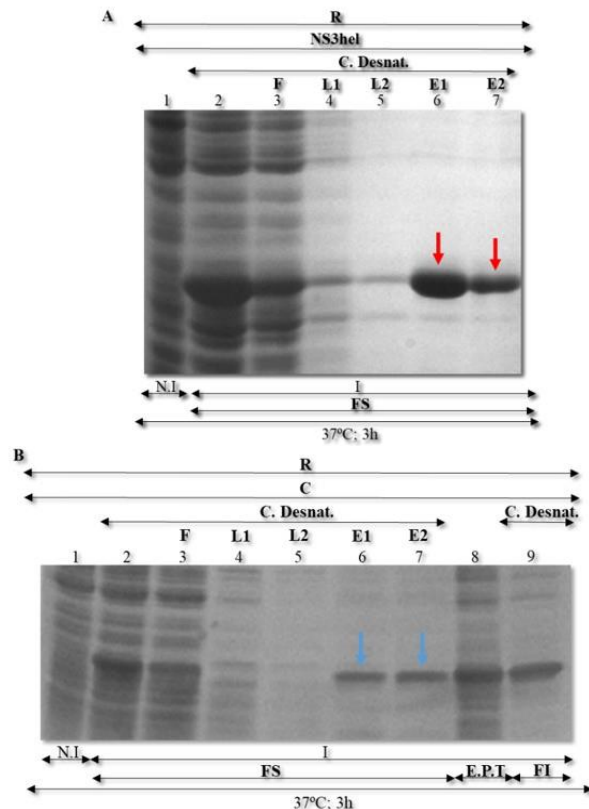
Primeiramente, a técnica foi realizada em condições desnaturantes apenas para verificação da presença de caudas de His<sub>6</sub> nas proteínas recombinantes produzidas, por esta abordagem ser metodologicamente mais expedita e conveniente. A desnaturação proteica foi assegurada pela presença de uma elevada concentração de ureia (7M). O processo de purificação foi realizado com recurso a um gradiente de pH, ou seja, para ligação da histidina ao níquel é necessário que a mesma se encontre desprotonada, para isso o tampão de lise/ligação utilizado possuía pH 8. Por outro lado, para a eluição da proteína, é essencial que as histidinas se encontrem protonadas, por isso, as proteínas foram eluídas com um tampão com pH 4,5.

Os extratos proteicos contendo NS3hel(ot) ou C(ot-del) produzidas em condição padrão foram sujeitos a purificação em condições desnaturantes (ver secção 2.9.7). Para as proteínas C e NS3hel, os filtrados correspondentes aos passos de fixação à coluna e de lavagens e eluatos foram analisados por SDS-PAGE (ver secção 2.9.5). Para a proteína NS3hel, a análise permitiu verificar a presença da mesma nas frações correspondentes aos eluatos, pelo que se pode confirmar a presença de caudas de His<sub>6</sub> na mesma. Ainda assim, nestas condições, pode observar-se perda de proteína heteróloga nas frações correspondentes aos filtrados, a qual pode ser explicada pela elevada concentração de proteína NS3hel no extrato e pela consequente insuficiente quantidade de íons níquel disponíveis na coluna para ligação com a mesma. Se o objetivo do estudo passasse pela purificação em condições desnaturantes, nestas condições, os eluatos revelaram uma



pequena quantidade de proteínas contaminantes, contudo pode-se concluir que NS3hel foi obtida com um nível de purificação considerável (Figura 3.15 A).

Curiosamente, para a proteína C, nestas condições, não se conseguiu verificar a sua presença nas frações correspondentes aos eluatos (resultados não apresentados). A resolução deste problema foi abordada através do ajustamento da composição dos tampões de lise, lavagem e eluição, tendo por objetivo o aumento da força iônica dos mesmos, o que foi conseguido com a adição de 500mM NaCl. Assim, após adição deste sal, foi possível a verificação da presença de caudas de hexa-histidina na proteína C-His<sub>6</sub>. Para além de aumentar a força iônica dos tampões, a presença de NaCl ajuda a reproduzir as condições físico-químicas mais próximas do ambiente fisiologicamente relevante e a remover contaminantes da resina de *Ni-NTA*, pela redução de ligações não específicas à mesma. Nestas condições, também se pode observar perda de proteína heteróloga nas frações correspondentes aos filtrados e insolúvel, esta última resultou de precipitação proteica durante o processo de lise celular. Os eluatos revelaram uma pequena quantidade de proteínas contaminantes, podendo concluir que C foi obtida com um nível de purificação considerável (Figura 3.15 B).



**Figura 3.15.** Análise, após separação por SDS-PAGE a 13% dos produtos resultantes da purificação por cromatografia de afinidade com íons níquel das proteínas NS3hel (A) e C (B) em condições desnaturantes.

**A e B** - 1 e 2 – extratos proteicos de culturas controlo e correspondente à fração solúvel (FS) de culturas onde foi induzida a produção heteróloga, respetivamente. Filtrados correspondentes aos passos de fixação (3) e de lavagens (4 e 5). 6 e 7 – eluatos. 8 e 9 – extrato de proteína solúvel total (E.P.T) e correspondente à fração insolúvel (FI), de culturas onde foi induzida a produção heteróloga, respetivamente. O tempo de indução e a temperatura a que esta ocorreu estão indicados na figura.

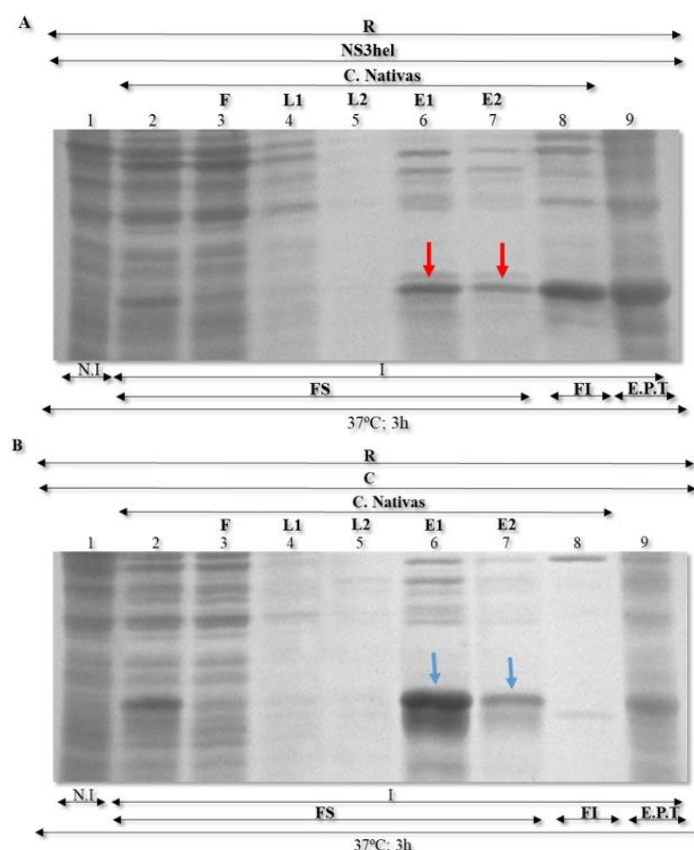
Após verificação da presença de caudas de His<sub>6</sub> nas proteínas recombinantes produzidas foi possível a purificação das mesmas em condições nativas, *i.e.*, na ausência de qualquer agente desnaturante. O princípio do processo é semelhante ao empregue em condições desnaturantes, contudo, a eluição é realizada pela utilização de 0,5M de imidazole. As histidinas contêm um anel de imidazole que lhes garante elevada afinidade pelos iões níquel presentes na coluna de purificação, assim se empregue uma elevada concentração de imidazole, o mesmo compete com as caudas de His<sub>6</sub> pela ligação aos iões, permitindo a purificação após passagem pela coluna de cromatografia.

Os extratos proteicos obtidos (correspondentes às frações solúveis enriquecidas em NS3hel e C) nas condições em que se obteve maior grau de solubilidade proteica (ver secção 3.4.2) foram sujeitos a purificação em condições nativas (ver secção 2.9.8). Os filtrados e eluatos obtidos foram analisados por SDS-PAGE (ver secção 2.9.5). A análise permitiu verificar a presença das proteínas NS3hel (Figura 3.16 A) e C (Figura 3.16 B) nas frações correspondentes aos eluatos. Apesar de todo o processo de otimização descrito, a proteína NS3hel foi obtida maioritariamente na fração insolúvel enquanto que C foi recuperada na sua totalidade na fração solúvel correspondente aos eluatos (Figura 3.16).

Face aos objetivos traçados, seria desejável a purificação das proteínas em condições nativas. Apesar de se ter obtido um elevado rendimento de produto sob a forma de CI, a purificação a partir dos mesmos e recuperação *in vitro* da solubilidade e bioatividade, depende de complexos processos de renaturação que não garantem elevados níveis de recuperação proteica, e representam custos elevados (Sørensen e Mortensen, 2005a). É de notar que a obtenção de uma elevada quantidade de proteína solúvel não significa qualidade conformacional, ou seja, a proteína adotou uma conformação solúvel estável, mas a arquitetura exata do centro ativo pode não ser a apropriada para a sua atividade (Rosano e Ceccarelli, 2014).

Nestas condições, os eluatos revelaram uma quantidade considerável de proteínas contaminantes (proteínas com múltiplos resíduos de histidina e/ou chaperones ligadas

diretamente à resina e/ou à proteína recombinante), apesar de terem sido utilizados 20mM e 30mM de imidazole nas soluções de ligação e lavagem, respetivamente, como estratégia de diminuição da ligação inespecífica. Ainda assim fica como prova de conceito o sucesso do processo de purificação em condições nativas. Em futuros estudos a otimização do protocolo de purificação em condições nativas será necessário, por exemplo, através de passos adicionais de purificação com cromatografia de permuta iónica e/ou de exclusão molecular (Gräslund *et al.*, 2008; Jain *et al.*, 2016).



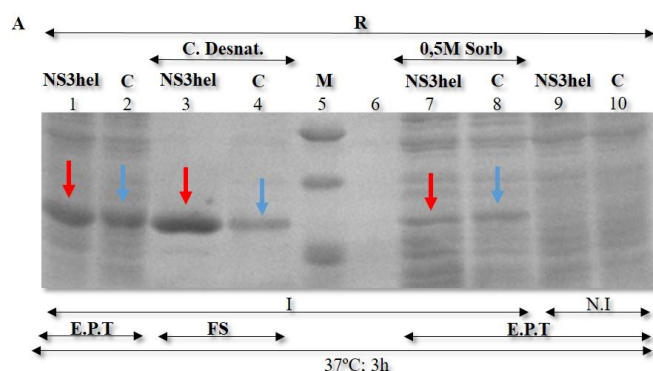
**Figura 3.16.** Análise, após separação por SDS-PAGE a 13%, dos produtos resultantes da purificação por cromatografia de afinidade com iões níquel das proteínas NS3hel (**A**) e C (**B**) em condições nativas. **A** e **B** - 1 e 2 – extratos proteicos de culturas controlo e correspondente à fração solúvel (FS) de culturas onde foi induzida a produção heteróloga, respetivamente. Filtrados correspondentes aos passos de fixação (3) e de lavagens (4 e 5). 6 e 7 – eluatos. 8 e 9 – extrato proteico correspondente à fração insolúvel (FI) e extrato de proteína solúvel total (E.P.T) de culturas onde foi induzida a produção heteróloga, respetivamente. O tempo de indução e a temperatura a que esta ocorreu estão indicados na figura.

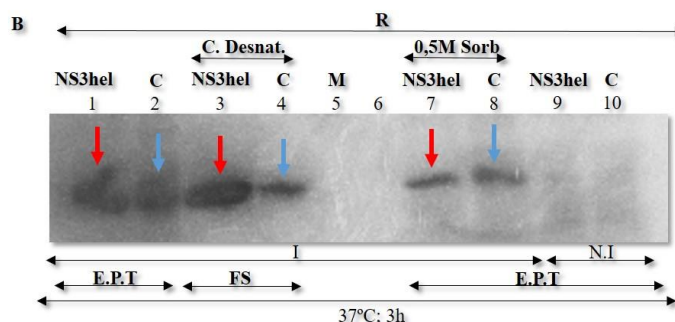
### 3.4.5. Verificação da presença de cauda de His<sub>6</sub>, por *western blot*, nas proteínas recombinantes produzidas

Em conjunto com o processo de purificação por cromatografia de afinidade, a presença de caudas de His<sub>6</sub> nas proteínas recombinantes produzidas foi adicionalmente

verificada utilizando o método de hibridação de anticorpos a proteínas transferidas para um suporte por *western blot*. Desta feita, esta abordagem permite a identificação do parceiro de fusão (His<sub>6</sub>) em proteínas imobilizadas num suporte sólido, utilizando anticorpos monoclonais anti-His<sub>6</sub> (ver secção 2.9.9).

O *western blot* pressupõe a separação das proteínas desnaturadas com base na respetiva massa molecular por SDS-PAGE, seguida da sua transferência para um suporte sólido (membrana), bloqueio da membrana com uma solução proteica para limitar a ligação inespecífica dos anticorpos à mesma, incubação com anticorpos primário e secundário e, por fim, a deteção de complexos antígeno/anticorpo primário/anticorpo secundário conjugado. A análise de imunodeteção referente a extratos proteicos totais de culturas bacterianas de *E. coli Rosetta(DE3)pLysS*, produtoras de NS3hel ou C (induzidas na condição padrão e na que obteve maior grau de solubilidade proteica), a eluatos resultantes do processo de purificação em condições desnaturantes e a extratos proteicos totais derivados de culturas não induzidas permitiu confirmar, uma vez mais, de forma específica, a presença de caudas de His<sub>6</sub> nas proteínas recombinantes produzidas (NS3hel e C com  $\approx$  15kDa e 13kDa, respetivamente), nomeadamente, nos extratos proteicos totais de culturas induzidas e nos eluatos resultantes do processo de purificação (Figura 3.17 B). Adicionalmente foi possível observar uma pequena quantidade de proteínas contaminantes, nomeadamente, proteínas com múltiplos resíduos de histidina, o que reflete um insuficiente bloqueio da membrana com solução proteica e/ou lavagem da membrana (Figura 3.17 B).





**Figura 3.17.** A) Análise, após separação por SDS-PAGE a 13%, de extratos proteicos totais de culturas induzidas, com (7 e 8) e sem (1 e 2) suplementação do meio LB líquido com 0,5M D-sorbitol, a 37°C durante 3h, de eluatos resultantes do processo de purificação em condições desnaturantes (3 e 4) e de extratos proteicos totais derivados de culturas controle (9 e 10). B) Detecção da presença de caudas de His<sub>6</sub> nas proteínas NS3hel e C por *western blot* (pistas equivalentes em A). 5 - M - marcador de proteínas com massas moleculares conhecidas (*Precision Plus Protein™ Dual Color Standards*).

Em suma foi possível a produção/purificação de antígenos virais de ZIKV em *E. coli*, nomeadamente, a proteína estrutural da cápside (C) e região C-terminal com atividade enzimática de RNAhel (NS3hel) de um provável constituinte do complexo replicativo (proteína NS3). Para a proteína C foi maximizada a sua produção sob a forma solúvel. Contudo, apesar de toda a otimização descrita de forma a melhorar a solubilidade proteica (ver secção 3.4.2), a proteína NS3hel foi expressa maioritariamente sob a forma insolúvel.

Neste estudo o péptido de fusão His<sub>6</sub> foi colocado em C-terminal (ver secção 3.2) para purificação das proteínas através de cromatografia de afinidade com iões níquel. No entanto, para além de pudermos ainda ser exploradas as possibilidades de indução a partir de uma miríade de outros vetores, uma simples alternativa, mantendo o vetor utilizado, passaria por explorar a possibilidade que este oferece de permitir a adição de um outro parceiro de fusão alternativo, designado *S-tag*, na região N-terminal das proteínas a produzir. Curiosamente, não só diferentes parceiros de fusão podem contribuir de forma diferencial para a expressão/solubilidade/estabilidade das proteínas a que se encontrem ligados, mas simplesmente a sua posição influencia os resultados obtidos (Costa *et al.*, 2014). Uma outra sugestão passa pela utilização de um parceiro de fusão que seja uma proteína altamente solúvel e estável, que claramente promova a solubilidade das proteínas alvo (MBP, NusA, GST, Trx, SUMO e Fh8). Curiosamente, no decurso deste trabalho, NS3hel de ZIKV foi produzida sob a forma solúvel através da colocação em N-terminal do parceiro de fusão, His<sub>6</sub>-SUMO (Jain *et al.*, 2016). Alternativamente, o aumento da solubilidade das proteínas obtidas pode ser tentativamente explorado por via da co-expressão de chaperones moleculares (GroEL-GroES, DnaK/DnaJ/GrpE e TF).

Tal como referido acima, a redução da taxa de síntese proteica através da diminuição da temperatura de incubação pode permitir o aumento da solubilidade proteica ao criar um ambiente não favorável à agregação proteica, contudo, neste estudo a abordagem não surtiu efeito. A redução da taxa de crescimento e de síntese proteica refletiu-se num baixo rendimento proteico total e além disso ocorreu uma diminuição na eficiência do mecanismo das chaperones moleculares a baixas temperaturas. Por isso, a expressão heteróloga pode ser realizada utilizando estirpes de expressão com chaperones adaptadas a baixas temperaturas, nomeadamente, a ArticExpress<sup>TM</sup> (Stratagene, E.U.A). Também podem ser utilizados sistemas nos quais o nível de expressão é ajustável (Tuner<sup>TM</sup> (Novagen)). Neste momento encontra-se descrita a estrutura em alta resolução da proteína NS3hel, a 1,62-Å (Jain *et al.*, 2016) e 1,8-Å (Tian *et al.*, 2016), pelo que se podia optar por um rastreio *in silico* de potenciais epítomos/determinantes de superfície de uma infeção *in vivo*, de forma a realizar diretamente a sua síntese *de novo* e produzir anticorpos específicos a partir dos mesmos.

Por último, coloca-se como opção alternativa, a utilização de um sistema diferente de expressão heteróloga de proteínas recombinantes, nomeadamente, as células de insetos (exemplos, Sf9 e Hi5), oferecendo entre as demais vantagens, a possibilidade de modificações pós-traducionais nas proteínas a ser produzidas, incluindo, fosforilação, N- e O- glicosilação, acilação e estabelecimento de ligação dissulfeto, aquisição de uma estrutura tridimensional correta, níveis elevados de expressão, fácil produção em larga escala e segurança (Demain e Vaishnav, 2009).

#### 4. CONCLUSÕES E PERSPETIVAS FUTURAS

O presente estudo veio permitir a produção/purificação de dois antígenos virais de ZIKV em *E. coli* *Rosetta(DE3)pLysS*, nomeadamente, a proteína estrutural da cápside (C) e região C-terminal de um provável constituinte do complexo replicativo (proteína NS3), correspondente ao domínio funcional desta enzima viral, ao qual está associada atividade de helicase de RNA (NS3hel). Para tal, as suas sequências codificantes de genótipo selvagem ou correspondentes a versões sintéticas com codões otimizados artificialmente, para se ajustarem à maquinaria de biossíntese do hospedeiro que viria a ser utilizado para a sua produção, foram clonadas no vetor de expressão (pET-29a), de modo a possibilitar a expressão em *E. coli* como proteínas de fusão com caudas de His<sub>6</sub> na extremidade C-terminal.

A análise, em paralelo, da expressão proteica de sequências virais de genótipo selvagem e versões sintéticas com codões otimizados nas estirpes de *E. coli*, *BL21(DE3)pLysS* e *Rosetta(DE3)pLysS* permitiu determinar que para NS3hel-His<sub>6</sub>, a otimização artificial da sua sequência nucleotídica possibilitou com elevada eficiência a produção desta proteína. Por outro lado, a produção da proteína C-His<sub>6</sub> foi conseguida não só pela otimização da expressão proteica por via da alteração da sua sequência codificante, mas também pela deleção de parte da sequência nucleotídica que codifica o extremo C-terminal (hidrofóbico) desta proteína, após avaliação de uma variedade de parâmetros que frequentemente determinam não só a abundância, mas também a solubilidade das proteínas produzidas em hospedeiros heterólogos (temperatura de crescimento, concentração do indutor, heterogeneidade da população bacteriana e cinética de produção). O ensaio de verificação da solubilidade das proteínas C-His<sub>6</sub> e NS3hel-His<sub>6</sub> permitiu verificar que a proteína NS3hel-His<sub>6</sub> apresenta-se maioritariamente na fração insolúvel, enquanto a proteína C-His<sub>6</sub> se distribuiu maioritariamente na fração solúvel. No presente estudo foram avaliados vários parâmetros de forma a melhorar a solubilidade das proteínas recombinantes produzidas, dos quais se destacou o uso do aditivo D-sorbitol no meio de crescimento, uma vez que, esta adição permitiu o aumento de produção das proteínas C-His<sub>6</sub> e NS3hel-His<sub>6</sub> sob a forma solúvel. Assim sendo, concluiu-se que as condições ótimas de solubilidade proteica ocorreram quando a indução foi efetuada com 1mM de IPTG em meio LB líquido suplementado com 0,5M D-sorbitol e incubação até uma D.O.<sub>600nm</sub>, de  $\approx 0,8$ , durante 3h, a 37°C com agitação (200-220rpm).

A identidade das proteínas de fusão C-His<sub>6</sub> ( $\approx$  13kDa) e NS3hel-His<sub>6</sub> ( $\approx$  15kDa) foi verificada por purificação, em condições desnaturantes, das mesmas por cromatografia de afinidade com íons níquel e por *western blot*, utilizando anticorpos monoclonais anti-His<sub>6</sub>. Se o objetivo do estudo passasse pela purificação em condições desnaturantes, nestas condições foi observada uma pequena quantidade de proteínas contaminantes, podendo concluir que NS3hel-His<sub>6</sub> e C-His<sub>6</sub> foram obtidas com um nível de purificação considerável. Igualmente foi possível a purificação das proteínas em estudo em condições nativas. Nestas condições foi observada uma quantidade considerável de proteínas contaminantes. Ainda assim fica como prova de conceito o sucesso do processo de purificação em condições nativas.

O presente trabalho deixa em aberto vários desafios, nomeadamente, a maximização da produção de proteína NS3hel-His<sub>6</sub> sob a forma solúvel, podendo esta ser realizada tendo em conta os caminhos sugeridos acima, dos quais se destaca, a utilização de um sistema eucariota de expressão heteróloga de proteínas recombinantes, nomeadamente, as células de insetos (exemplos, Sf9 e Hi5). Este tipo de sistema oferece entre as demais vantagens, uma mudança no microambiente celular de produção heteróloga em relação a *E. coli*, que pode ser traduzida numa diferente abundância e tipo de chaperones moleculares existentes, resultando num diferente nível de expressão heteróloga e solubilidade das proteínas pretendidas. Em alternativa pode ser realizada a otimização do protocolo de renaturação *in vitro* de NS3hel-His<sub>6</sub>.

Em futuros estudos, a otimização do protocolo de purificação em condições nativas também será um passo necessário. As proteínas virais recombinantes obtidas podem vir a ser utilizadas para a produção de anticorpos específicos, os quais, por sua vez, poderão ser empregues em estudos de estratégia replicativa em células permissivas. Por outro lado, pode ser feita a determinação da estrutura tridimensional das proteínas virais recombinantes após cristalização. Além de estudos funcionais, por exemplo, testando a ligação de C ao RNA viral ou da atividade de helicase de NS3hel-His<sub>6</sub> (ou mesmo após eliminação do companheiro de fusão), visando o rastreio, *in vitro*, de compostos que a possa especificamente inibir, e que possam ser explorados no sentido da produção de fármacos antivirais.



## 5. BIBLIOGRAFIA

Adams Waldorf, K. M., J. E. Stencel-Baerenwald, R. P. Kapur, C. Studholme, E. Boldenow, J. Vornhagen *et al.* 2016. Fetal brain lesions after subcutaneous inoculation of Zika virus in a pregnant nonhuman primate. *Nat Med.* 22:1256–59.

Aubry, M., A. Teissier, M. Huart, S. Merceron, J. Vanhomwegen, C. Roche *et al.* 2017. Zika virus seroprevalence, French Polynesia, 2014–2015. *Emerg. Infect. Dis.* 23: 669–72.

Baneyx, F. 1999. Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* 10:411-21.

Baneyx, F., M. Mujacic. 2004. Recombinant protein folding and misfolding in *Escherichia coli*. *Nat. Biotechnol.* 22:1399-408.

Bardina, S. V., P. Bunduc, S. Tripathi, J. Duehr, J. J. Frere, J. A. Brown *et al.* 2017. Enhancement of Zika virus pathogenesis by preexisting ant flavivirus immunity. *Science.* 356:175-80.

Baronti, C., G. Piorkowski, R. N. Charrel, L. Boubis, I. Leparç-Goffart, X. de Lamballerie. 2014. Complete coding sequence of Zika virus from a French Polynesia outbreak in 2013. *Genome Announc.* 2:e00500–14.

Barzon, L., M. Pacenti, A. Berto, A. Sinigaglia, E. Franchin, E. Lavezzo *et al.* 2016a. Isolation of infectious Zika virus from saliva and prolonged viral RNA shedding in a traveller returning from the Dominican Republic to Italy, January 2016. *Euro Surveill.* 21:30159.

Barzon, L., M. Trevisan, A. Sinigaglia, E. Lavezzo, G. Palù. 2016b. Zika virus: from pathogenesis to disease control. *FEMS Microbiol. Lett.* 363. pii:fnw202.

Baud, D., D. J. Gubler, B. Schaub, M. C. Lanteri, D. Musso. 2017. An update on Zika virus infection. *Lancet.* pii: S0140-6736(17)31450-2.

Bayer, A., N. J. Lennemann, Y. Ouyang, J. C. Bramley, S. Morosky, E. T. Marques Jr. *et al.* 2016. Type III interferons produced by human placental trophoblasts confer protection against Zika virus infection. *Cell Host Microbe.* 19:705–12.

Besnard, M., S. Lastere, A. Teissier, V. M. Cao-Lormeau, D. Musso. 2014. Evidence of perinatal transmission of Zika virus, French Polynesia, December 2013 and February 2014. *Euro Surveill.* 19. pii: 20751.

Birnboim, H. C., J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic. Acids. Res.* 7:1513-23.

Blitvich, B. J., A. E. Firth. 2017. A Review of Flaviviruses that Have No Known Arthropod Vector. *Viruses.* 9. pii: E154.

- Blommel, P. G., K. J. Becker, P. Duvnjak, B. G. Fox. 2007. Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition. *Biotechnol. Prog.* 23:585–98.
- Bornhorst, J. A., J. J. Falke. 2000. Purification of proteins using polyhistidine affinity tags. *Methods Enzymol.* 326:245–54.
- Brasil, P., G. A. Calvet, A. M. Siqueira, M. Wakimoto, P. C. de Sequeira, A. Nobre *et al.* 2016. Zika virus outbreak in Rio de Janeiro, Brazil: clinical characterization, epidemiological and virological aspects. *PLoS Negl. Trop. Dis.* 10: e0004636.
- Brito, C. A., M. T. Cordeiro. 2016. One year after the Zika virus outbreak in Brazil: from hypotheses to evidence. *Rev. Soc. Bras. Med. Trop.* 49: 537–43.
- Calvet, G., R. S. Aguiar, A. S. Melo, S. A. Sampaio, I. de Filippis, A. Fabri *et al.* 2016. Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. *Lancet Infect. Dis.* 16:653–60.
- Cardoso, C. W., I. A. Paploski, M. Kikuti, M. S. Rodrigues, M. M. Silva, G. S. Campos *et al.* 2015. Outbreak of exanthematous illness associated with Zika, chikungunya, and dengue viruses, Salvador, Brazil. *Emerg. Infect. Dis.* 21:2274–6.
- Carrió, M. M., A. Villaverde. 2001. Protein aggregation as bacterial inclusion bodies is reversible. *FEBS Lett.* 489:29–33.
- Cauchemez, S., M. Besnard, P. Bompard, T. Dub, P. Guillemette-Artur, D. Eyrolle-Guignot *et al.* 2016. Association between Zika virus and microcephaly in French Polynesia, 2013–15: a retrospective study. *Lancet.* 387:2125–32.
- Chae, Y. K., S. H. Kim, J. L. Markley. 2017. Relationship between recombinant protein expression and host metabolome as determined by two-dimensional NMR spectroscopy. *PLoS One.* 12:e0177233.
- Chan, J. F., A. J. Zhang, C. C. Chan, C. C. Yip, W. W. Mak, H. Zhu *et al.* 2016. Zika virus infection in dexamethasone-immunosuppressed mice demonstrating disseminated infection with multi-organ involvement including orchitis effectively treated by recombinant type I interferons. *EBioMedicine.* 14: 112–22.
- Chen, J. C., Z. Wang, H. Huang, S. H. Weitz, A. Wang, X. Qiu *et al.* 2016. Infection of human uterine fibroblasts by Zika virus *in vitro*: implications for viral transmission in women. *Int. J. Infect. Dis.* 51:139–40.
- Chung, C. T., S. L. Niemela, R. H. Miller. 1989. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. U S A.* 7:2172–5.
- Coffey, L. L., P. A. Pesavento, R. I. Keesler, A. Singapuri, J. Watanabe, R. Watanabe *et al.* 2017. Zika virus tissue and blood compartmentalization in acute infection of rhesus macaques. *PLoS One.* 12:e0171148.

- Costa, S., A. Almeida, A. Castro, L. Domingues. 2014. Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Front. Microbiol.* 5:63.
- Creange, A. 2016. Guillain–Barré syndrome: 100 years on. *Rev. Neurol.* 172:770–4.
- Dang, J., S. K. Tiwari, G. Lichinchi, Y. Qin, V. S. Patil, A. M. Eroshkin *et al.* 2016. Zika virus depletes neural progenitors in human cerebral organoids through activation of the innate immune receptor TLR3. *Cell Stem Cell.* 19:258–65.
- Davidson, A., S. Slavinski, K. Komoto, J. Rakeman, D. Weiss. 2016. Suspected female-to-male sexual transmission of Zika virus - New York City, 2016. *MMWR Morb. Mortal. Wkly. Rep.* 65:716–7.
- de Marco, A., L. Vigh, S. Diamant, P. Goloubinoff. 2005. Native folding of aggregation-prone recombinant proteins in *Escherichia coli* by osmolytes, plasmid- or benzyl alcohol-overexpressed molecular chaperones. *Cell Stress Chaperones.* 10:329–39.
- de Marco, A., E. Deuerling, A. Mogk, T. Tomoyasu, B. Bukau. 2007. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnol.* 7:32.
- Deckard, D. T., W. M. Chung, J. T. Brooks, J. C. Smith, S. Woldai, M. Hennessey *et al.* 2016. Male-to-Male Sexual Transmission of Zika Virus-Texas, January 2016. *MMWR Morb. Mortal. Wkly. Rep.* 65:372–4.
- Demain, A. L., P. Vaishnav. 2009. Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* 27:297–306.
- Derman, A. I., W. A. Prinz, D. Belin, J. Beckwith. 1993. Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science.* 262:1744–7.
- Dick, G. W., S. F. Kitchen, A. J. Haddow. 1952. Zika virus. I. Isolations and serological specificity. *Trans. R. Soc. Trop. Med. Hyg.* 46:509–20.
- Doherty, A. J., B. A. Connolly, A. F. Worrall. 1993. Overproduction of the toxic protein, bovine pancreatic DNaseI, in *Escherichia coli* using a tightly controlled T7-promoter-based vector. *Gene.* 136:337–40.
- Dong, H., L. Nilsson, C. G. Kurland. 1995. Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *J. Bacteriol.* 177:1497–504.
- Dong, H., L. Nilsson, C. G. Kurland. 1996. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.* 260:649–63.
- Dubendorff, J. W., F. W. Studier. 1991. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* 219:45–59.

- Duffy, M. R., T. H. Chen, W. T. Hancock, A. M. Powers, J. L. Kool, R. S. Lanciotti *et al.* 2009. Zika virus outbreak on Yap Island, Federated States of Micronesia. *N. Engl. J. Med.* 360:2536–43.
- Dupont-Rouzeyrol, M., O. O'Connor, E. Calvez, M. Daurès, M. John, J. P. Grangeon *et al.* 2015. Co-infection with Zika and dengue viruses in 2 patients, New Caledonia, 2014. *Emerg. Infect. Dis.* 21:381-2.
- Ellison, D. W., J. T. Ladner, R. Buathong, M. T. Alera, M. R. Wiley, L. Hermann *et al.* 2016. Complete genome sequences of Zika virus strains isolated from the blood of patients in Thailand in 2014 and the Philippines in 2012. *Genome Announc.* 4:e00359-16.
- European Centre for Disease Prevention and Control. 2017. *Aedes aegypti* - current known distribution in Europe, April 2017. European Centre for Disease Prevention and Control, Solna, Suécia, 22 de maio de 2017. Disponível em: <https://ecdc.europa.eu/en/publications-data/aedes-aegypti-current-known-distribution-europe-april-2017>. [Acedido: 24 de julho de 2017].
- Faria, N. R., R. D. Azevedo, M. U. Kraemer, R. Souza, M. S. Cunha, S. C. Hill *et al.* 2016. Zika virus in the Americas: early epidemiological and genetic findings. *Science.* 352:345-9.
- Faria, N. R., J. Quick, I. M. Claro, J. Thézé, J. G. de Jesus, M. Giovanetti *et al.* 2017. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature.* 546:406-10.
- Faye, O., C. C. Freire, A. Iamarino, O. Faye, J. V. de Oliveira, M. Diallo *et al.* 2014. Molecular evolution of Zika virus during its emergence in the 20th century. *PLoS Negl. Trop. Dis.* 8:e2636.
- Fernandez-Castane, A., C. E. Vine, G. Caminal, J. Lopez-Santin. 2012. Evidencing the role of lactose permease in IPTG uptake by *Escherichia coli* in fed-batch high cell density cultures. *J. Biotechnol.* 157:391–8.
- Ferrer, M., H. Lunsdorf, T. N. Chernikova, M. Yakimov, K. N. Timmis, P. N. Golyshin. 2004. Functional consequences of single:double ring transitions in chaperonins: life in the cold. *Mol. Microbiol.* 53:167–82.
- Foy, B. D., K. C. Kobylinski, J. L. Chilson Foy, B. J. Blitvich, A. Travassos da Rosa, A. D. Haddow *et al.* 2011. Probable non-vector-borne transmission of Zika virus, Colorado, USA. *Emerg. Infect. Dis.* 17:880-2.
- Galloway, C. A., M. P. Sowden, H. C. Smith. 2003. Increasing the yield of soluble recombinant protein expressed in *E. coli* by induction during late log phase. *Biotechniques.* 34:524-6, 528, 530.
- Garcez, P. P., E. C. Loiola, R. Madeiro da Costa, L. M. Higa, P. Trindade, R. Delvecchio *et al.* 2016. Zika virus impairs growth in human neurospheres and brain organoids. *Science.* 352:816-8.

- Govero, J., P. Esakky, S. M. Scheaffer, E. Fernandez, A. Drury, D. J. Platt *et al.* 2016. Zika virus infection damages the testes in mice. *Nature*. 540:438–42.
- Grard, G., M. Caron, I. M. Mombo, D. Nkoghe, S. M. Ondo, D. Jiolle *et al.* 2014. Zika virus in Gabon (Central Africa) -2007: a new threat from *Aedes albopictus*?. *PLoS Negl. Trop. Dis.* 8:e2681.
- Gräslund, S., P. Nordlund, J. Weigelt, J. Bray, O. Gileadi, S. Knapp. 2008. Protein production and purification. *Nat. Methods*. 5:135–46.
- Grodberg, J., J. J. Dunn. 1988. ompT encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *J. Bacteriol.* 170: 1245–53.
- Guerbois, M., I. Fernandez-Salas, S. R. Azar, R. Danis-Lozano, C. M. Alpuche-Aranda, G. Leal *et al.* 2016. Outbreak of Zika Virus Infection, Chiapas State, Mexico, 2015, and First Confirmed Transmission by *Aedes aegypti* Mosquitoes in the Americas. *J. Infect. Dis.* 214:1349–56.
- Gustafsson, C., S. Govindarajan, J. Minshull. 2004. Codon bias and heterologous protein expression. *Trends Biotechnol.* 22:346–53.
- Haddow, A. D., A. J. Schuh, C. Y. Yasuda, M. R. Kasper, V. Heang, R. Huy *et al.* 2012. Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. *PLoS Negl. Trop. Dis.* 6:e1477.
- Haider, S. R., H. J. Reid, B. L. Sharp. 2010. Modification of tricine-SDS-PAGE for online and offline analysis of phosphoproteins by ICP-MS. *Anal. Bioanal. Chem.* 397:655–64.
- Hamel, R., O. Dejarnac, S. Wichit, P. Ekchariyawat, A. Neyret, N. Luplertlop *et al.* 2015. Biology of Zika virus infection in human skin cells. *J. Virol.* 89:8880–96.
- Hammarstrom, M., N. Hellgren, S. Van Den Berg, H. Berglund, T. Hard. 2002. Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Sci.* 11:313–21.
- Hartley, D. L., J. F. Kane. 1988. Properties of inclusion bodies from recombinant *Escherichia coli*. *Biochem. Soc. Trans.* 16:101–2.
- He, Y., K. Wang, N. Yan. 2014. The recombinant expression systems for structure determination of eukaryotic membrane proteins. *Protein Cell.* 5:658–72.
- Hellmuth, K., D. J. Korz, E. A. Sanders, W. D. Deckwer. 1994. Effect of growth rate on stability and gene expression of recombinant plasmids during continuous and high cell density cultivation of *Escherichia coli* TG1. *J. Biotechnol.* 32:289–98.
- Hennessey, M., M. Fischer, J. E. Staples. 2016. Zika virus spreads to new areas-region of the Americas, May 2015–January 2016. *MMWR Morb. Mortal. Wkly. Rep.* 65:55–8.
- Jain, R., J. Coloma, A. García-Sastre, A. K. Aggarwal. 2016. Structure of NS3 helicase from Zika virus. *Nat. Struct. Mol. Biol.* 23: 752–4.

- Kaufmann, B., M. G. Rossmann. 2011. Molecular mechanisms involved in the early steps of flavivirus cell entry. *Microbes Infect.* 13:1–9.
- Keasling, J. D. 1999. Gene-expression tools for the metabolic engineering of bacteria. *Trends Biotechnol.* 17:452-60.
- Kleber de Oliveira, W., J. Cortez-Escalante, W. T. De Oliveira, G. M. do Carmo, C. M. Henriques, G. E. Coelho *et al.* 2016. Increase in Reported Prevalence of Microcephaly in Infants Born to Women Living in Areas with Confirmed Zika Virus Transmission During the First Trimester of Pregnancy - Brazil, 2015. *MMWR Morb. Mortal. Wkly. Rep.* 65:242-7.
- Kostyuchenko, V. A., E. X. Lim, S. Zhang, G. Fibriansah, T. S. Ng, J. S. Ooi *et al.* 2016. Structure of the thermally stable Zika virus. *Nature.* 533:425–8.
- Kraemer, M. U., M. E. Sinka, K. A. Duda, A. Q. Mylne, F. M. Shearer, C. M. Barker *et al.* 2015. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *Elife.* 4:e08347.
- Kuehnert, M. J., S. V. Basavaraju, R. R. Moseley, L. L. Pate, S. A. Galel, P. C. Williamson *et al.* 2016. Screening of Blood Donations for Zika Virus Infection - Puerto Rico, April 3-June 11, 2016. *MMWR Morb. Mortal. Wkly. Rep.* 65:627-8.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.* 227: 680-5.
- Lanciotti, R. S., A. J. Lambert, M. Holodniy, S. Saavedra, C. Signor Ldel. 2016. Phylogeny of Zika virus in western hemisphere, 2015. *Emerg. Infect. Dis.* 22:933–5.
- Lazear, H. M., M. S. Diamond. 2016. Zika Virus: New Clinical Syndromes and Its Emergence in the Western Hemisphere. *J. Virol.* 90:4864-75.
- Lei, J., G. Hansen, C. Nitsche, C. D. Klein, L. Zhang, R. Hilgenfeld. 2016. Crystal structure of Zika virus NS2B-NS3 protease in complex with a boronate inhibitor. *Science.* 353:503-5.
- Lindenbach, B. D., H. J. Thiel, C. M. Rice. 2007. *Flaviviridae: The Viruses and Their Replication* In: Knipe, D. M., P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman *et al.* (Eds.). *Fields Virology.* 5<sup>th</sup> ed. Lippincott Williams & Wilkins Publishers, Philadelphia, PA. pp. 1101–52.
- Lindenbach, B. D., C. L. Murray, H. J. Thiel, C. M. Rice. 2013. *Flaviviridae.* In: Knipe, D. M., P. M. Howley, J. I. Cohen, D. E. Griffin, R. A. Lamb, M. A. Martin *et al.* (Eds.). *Fields Virology.* 6<sup>th</sup> ed. Lippincott Williams & Wilkins Publishers, Philadelphia, PA. pp. 712–46.
- Liu, L., W. Wu, X. Zhao, Y. Xiong, S. Zhang, X. Liu *et al.* 2016. Complete genome sequence of Zika virus from the first imported case in mainland China. *Genome Announc.* 4:e00291-16.

- Lockwood, C. J., S. T. Romero, K. Nielsen-Saines. 2017. Zika virus infection: Evaluation and management of pregnant women. UpToDate. Disponível em: <http://www.uptodate.com/contents/zika-virus-infection-evaluation-and-management-of-pregnant-women>. [Acedido: 29 de julho de 2017].
- Losen, M., B. Frolich, M. Pohl, J. Buchs. 2004. Effect of oxygen limitation and medium composition on *Escherichia coli* fermentation in shake-flask cultures. *Biotechnol. Prog.* 20:1062–8.
- Ma, W., S. Li, S. Ma, L. Jia, F. Zhang, Y. Zhang *et al.* Gao. 2016. Zika virus causes testis damage and leads to male infertility in mice. *Cell*. 167:1511–24.
- Makrides, S. C. 1996. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* 60:512–38.
- Marschall, L., P. Sagmeister, C. Herwig. 2016. Tunable recombinant protein expression in *E. coli*: enabler for continuous processing?. *Appl. Microbiol. Biotechnol.* 100:5719–28.
- Mergulhão, F. J., D. K. Summers, G. A. Monteiro. 2005. Recombinant protein secretion in *Escherichia coli*. *Biotechnol. Adv.* 23:177–202.
- Messens, J., J. F. Collet. 2006. Pathways of disulfide bond formation in *Escherichia coli*. *Int. J. Biochem. Cell Biol.* 38:1050–62.
- Miao, F., S. K. Drake, D. S. Kompala. 1993. Characterization of gene expression in recombinant *Escherichia coli* cells infected with phage lambda. *Biotechnol. Prog.* 9:153–9.
- Microcephaly Epidemic Research Group. 2016. Microcephaly in infants, Pernambuco State, Brazil, 2015. *Emerg. Infect. Dis.* 22:1090–3.
- Miroux, B., J. E. Walker. 1996. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* 260:289–98.
- Mlakar, J., M. Korva, N. Tul, M. Popović, M. Poljšak-Prijatelj, J. Mraz *et al.* 2016. Zika Virus Associated with Microcephaly. *N. Engl. J. Med.* 374:951–8.
- Moffatt, B. A., F. W. Studier. 1987. T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell*. 49:221–7.
- Montigny, C., F. Penin, C. Lethias, P. Falson. 2004. Overcoming the toxicity of membrane peptide expression in bacteria by upstream insertion of Asp-Pro sequence. *Biochim. Biophys. Acta.* 1660:53–65.
- Muller, D. A., P. R. Young. 2013. The flavivirus NS1 protein: molecular and structural biology, immunology, role in pathogenesis and application as a diagnostic biomarker. *Antiviral Res.* 98:192–208.

- Musso, D., T. Nhan, E. Robin, C. Roche, D. Bierlaire, K. Zisou *et al.* 2014a. Potential for Zika virus transmission through blood transfusion demonstrated during an outbreak in French Polynesia, November 2013 to February 2014. *Euro Surveill.* 19. pii:20771.
- Musso, D., E. J. Nilles, V. M. Cao-Lormeau. 2014b. Rapid spread of emerging Zika virus in the Pacific area. *Clin. Microbiol. Infect.* 20:O595-6.
- Musso, D., D. J. Gubler. 2016. Zika Virus. *Clin. Microbiol. Rev.* 29: 487–524.
- Nishihara, K., M. Kanemori, M. Kitagawa, H. Yanagi, T. Yura. 1998. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl. Environ. Microbiol.* 64:1694–9.
- Nishihara, K., M. Kanemori, H. Yanagi, T. Yura. 2000. Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. *Appl. Environ. Microbiol.* 66:884–9.
- Novy, R., B. Morris. 2001. Use of glucose to control basal expression in the pET System. *Innovations.* 13: 8–10.
- Novy, R., D. Drott, K. Yaeger, R. Mierendorf. 2001. Overcoming the codon bias of *E. coli* for enhanced protein expression. *Innovations.* 12:1–3.
- Onorati, M., Z. Li, F. Liu, A. M. Sousa, N. Nakagawa, M. Li *et al.* 2016. Zika Virus Disrupts Phospho-TBK1 Localization and Mitosis in Human Neuroepithelial Stem Cells and Radial Glia. *Cell Rep.* 16:2576-92.
- Pan American Health Organization. 2015. Epidemiological alert. Neurological syndrome, congenital malformations, and Zika virus infections. Implications for public health in the Americas. Pan American Health Organization, Washington, DC.
- Perera, R., M. Khaliq, R. J. Kuhn. 2008. Closing the door on flaviviruses: entry as a target for antiviral drug design. *Antiviral Res.* 80:11–22.
- Qian, X., H. N. Nguyen, M. M. Song, C. Hadiono, S. C. Ogden, C. Hammack *et al.* 2016. Brain-Region-Specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure. *Cell.* 165:1238-54.
- Quicke, K. M., J. R. Bowen, E. L. Johnson, C. E. McDonald, H. Ma, J. T. O'Neal *et al.* 2016. Zika virus infects human placental macrophages. *Cell Host Microbe.* 20:83–90.
- Rasmussen, S. A., D. J. Jamieson, M. A. Honein, L. R. Petersen. 2016. Zika Virus and Birth Defects--Reviewing the Evidence for Causality. *N. Engl. J. Med.* 374:1981-7.
- Rodríguez-Carmona, E., O. Cano-Garrido, M. Dragosits, M. Maurer, A. Mader, R. Kunert *et al.* 2012. Recombinant Fab expression and secretion in *Escherichia coli* continuous culture at medium cell densities: Influence of temperature. *Process Biochem.* 47: 446–52.



- Rosano, G. L., E. A. Ceccarelli. 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* 5:172.
- Roundy, C. M., S. R. Azar, S. L. Rossi, S. C. Weaver, N. Vasilakis. 2017. Insect-Specific Viruses: A Historical Overview and Recent Developments. *Adv. Virus. Res.* 98:119-46.
- Rozé, B., F. Najioullah, J. L. Fergé, K. Apetse, Y. Brouste, R. Cesaire *et al.* 2016. Zika virus detection in urine from patients with Guillain-Barré syndrome on Martinique, January 2016. *Euro Surveill.* 21:30154.
- Saiz, J. C., Á. Vázquez-Calvo, A. B. Blázquez, T. Merino-Ramos, E. Escribano-Romero, M. A. Martín-Acebes. 2016. Zika Virus: the Latest Newcomer. *Front. Microbiol.* 7:496.
- Sambrook, J., E. F. Fritsch, T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*. 2<sup>nd</sup> ed. Cold Spring Harbor Laboratory Press, Nova York.
- Sampath A., R. Padmanabhan. 2009. Molecular targets for flavivirus drug discovery. *Antiviral Res.* 81:6–15.
- Samsa, M. M., J. A. Mondotte, N. G. Iglesias, I. Assunção-Miranda, G. Barbosa-Lima, A. T. Da Poian *et al.* 2009. Dengue Virus Capsid Protein Usurps Lipid Droplets for Viral Particle Formation. *PLoS Pathog.* 5: e1000632.
- Sanger, F., S. Nicklen, A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U S A.* 74:5463–7.
- San-Miguel, T., P. Pérez-Bermúdez, I. Gavidia. 2013. Production of soluble eukaryotic recombinant proteins in *E. coli* is favoured in early log-phase cultures induced at low temperature. *Springerplus.* 2:89.
- Sardi, S., S. Somasekar, S. N. Naccache, A. C. Bandeira, L. B. Tauro, G. S. Campos *et al.* 2016. Coinfections from Zika and Chikungunya virus in Bahia, Brazil identified by metagenomic next-generation sequencing. *J. Clin. Microbiol.* 54:2348-53.
- Simmonds, P., P. Becher, J. Bukh, E. A. Gould, G. Meyers, T. Monath *et al.* 2017. ICTV Virus Taxonomy Profile: *Flaviviridae*. *J. Gen. Virol.* 98:2-3.
- Simpson, D. I. 1964. Zika virus infection in man. *Trans. R. Soc. Trop. Med. Hyg.* 58: 335–8.
- Sirohi, D., Z. Chen, L. Sun, T. Klose, T. C. Pierson, M. G. Rossmann *et al.* 2016. The 3.8 Å resolution cryo-EM structure of Zika virus. *Science.* 352:467–70.
- Smit, J. M., B. Moesker, I. Rodenhuis-Zybert, J. Wilschut. 2011. *Flavivirus* Cell Entry and Membrane Fusion. *Viruses.* 3: 160–71.
- Song, B. H., S. I. Yun, M. Woolley, Y. M. Lee. 2017. Zika virus: History, epidemiology, transmission, and clinical presentation. *J. Neuroimmunol.* 308:50-64.
- Song, H., J. Qi, J. Haywood, Y. Shi, G. F. Gao. 2016. Zika virus NS1 structure reveals diversity of electrostatic surfaces among flaviviruses. *Nat. Struct. Mol. Biol.* 23:456-8.

- Sørensen, H. P. e K. K. Mortensen. 2005a. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fact.* 4:1.
- Sørensen, H. P., K. K. Mortensen. 2005b. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* 115:113-28.
- Stewart, E. J., F. Aslund, J. Beckwith. 1998. Disulfide bond formation in the *Escherichia coli* cytoplasm: an *in vivo* role reversal for the thioredoxins. *EMBO J.* 17:5543–50.
- Studier, F. W. 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* 219:37-44.
- Studier, F. W., B. A. Moffatt. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189:113-30.
- Studier, F. W. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* 41:207-34.
- Tabata, T., M. Petitt, H. Puerta-Guardo, D. Michlmayr, C. Wang, J. Fang-Hoover *et al.* 2016. Zika virus targets different primary human placental cells, suggesting two routes for vertical transmission. *Cell Host Microbe.* 20:155–66.
- Tang, H., C. Hammack, S. C. Ogden, Z. Wen, X. Qian, Y. Li *et al.* 2016. Zika Virus Infects Human Cortical Neural Progenitors and Attenuates Their Growth. *Cell Stem Cell.* 18:587-90.
- Terpe, K. 2006. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 72:211-22.
- Tian, H., X. Ji, X. Yang, W. Xie, K. Yang, C. Chen *et al.* 2016. The crystal structure of Zika virus helicase: basis for antiviral drug design. *Protein Cell.* 7:450-4.
- van der Eijk, A. A., P. J. van Genderen, R. M. Verdijk, C. B. Reusken, R. Mögling, J. J. van Kampen *et al.* 2016. Miscarriage Associated with Zika Virus Infection. *N. Engl. J. Med.* 375:1002-4.
- Vera, A., N. Gonzalez-Montalban, A. Aris, A. Villaverde. 2007. The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnol. Bioeng.* 96:1101–6.
- Vouga, M., D. Baud. 2016. Imaging of congenital Zika virus infection: the route to identification of prognostic factors. *Prenat. Diagn.* 36:799–811.
- Wagner, S., M. M. Klepsch, S. Schlegel, A. Appel, R. Draheim, M. Tarry *et al.* 2008. Tuning *Escherichia coli* for membrane protein overexpression. *Proc. Natl. Acad. Sci. U.S.A.* 105:14371–6.
- Weaver, S.C., F. Costa, M. A. Garcia-Blanco, A. I. Ko, G. S. Ribeiro, G. Saade *et al.* 2016. Zika virus: History, emergence, biology, and prospects for control. *Antiviral Res.* 130:69-80.

Wikan, N., D. R. Smith. 2016. Zika virus: history of a newly emerging arbovirus. *Lancet Infect. Dis.* 16:e119-26.

World Health Organization. 1967. Arboviruses and human disease. WHO Tech. Rep. Ser. 369:1–84.

World Health Organization. 2016a. WHO Director-General Summarizes the Outcome of the Emergency Committee Regarding Clusters of Microcephaly and Guillain-Barré Syndrome. World Health Organization, Genebra, Suíça, 1 de fevereiro de 2016. Disponível em: <http://www.who.int/mediacentre/news/statements/2016/emergency-committee-zika-microcephaly/en/>. [Acedido: 13 de julho de 2017].

World Health Organization. 2016b. WHO Statement: Fifth Meeting of the Emergency Committee under the International Health Regulations (2005) Regarding Microcephaly, Other Neurological Disorders and Zika Virus. World Health Organization, Genebra, Suíça, 18 de novembro de 2016. Disponível em: <http://www.who.int/mediacentre/news/statements/2016/zika-fifth-ec/en/>. [Acedido: 13 de julho de 2017].

World Health Organization. 2017. Situation report. Zika virus microcephaly Guillain-Barré syndrome. World Health Organization, Genebra, Suíça, 10 de março de 2017. Disponível em: <http://reliefweb.int/report/world/zika-virus-microcephaly-and-guillain-barr-syndrome-situation-report-10-march-2017>. [Acedido: 13 de julho de 2017].

Yamaguchi, H., M. Miyazaki. 2014. Refolding Techniques for Recovering Biologically Active Recombinant Proteins from Inclusion Bodies. *Biomolecules*. 4: 235–51.

Yang, X., Y. Zhang. 2013. Effect of temperature and sorbitol in improving the solubility of carboxylesterases protein CpCE-1 from *Cydia pomonella* and biochemical characterization. *Appl. Microbiol. Biotechnol.* 97:10423-33.

Yin, J., G. Li, X. Ren, G. Herrler. 2007. Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *J. Biotechnol.* 127:335-47.

Yun, S. I., B. H. Song, J. C. Frank, J. G. Julander, I. A. Polejaeva, C. J. Davies *et al.* 2016. Complete Genome Sequences of Three Historically Important, Spatiotemporally Distinct, and Genetically Divergent Strains of Zika Virus: MR-766, P6-740, and PRVABC-59. *Genome Announc.* 4. pii: e00800–16.

Zanluca, C., V. C. Melo, A. L. Mosimann, G. I. Santos, C. N. Santos, K. Luz. 2015. First report of autochthonous transmission of Zika virus in Brazil. *Mem. Inst. Oswaldo Cruz.* 110:569–72.

Zou, J., X. Xie, Q. Y. Wang, H. Dong, M. Y. Lee, C. Kang *et al.* 2015. Characterization of dengue virus NS4A and NS4B protein interaction. *J. Virol.* 89:3455-70.

## 6. ANEXOS

## ANEXO I

**Tabela Anexo I.** Informação relativa às sequências nucleotídicas sintéticas otimizadas, as quais codificam C e NS3hel, respetivamente.

Região do genoma	Sequência nucleotídica sintética otimizada (5' → 3')
C	ATGAAAAACCCGAAAAAAAAAGAGTGGCGGTTTCCGCATCGTGAATAT GCTGAAACGTGGTGTTGCCCGTGTTAATCCGCTTGGTGGTCTTAAGCG TTTGCCGGCGGGTCTTTTGCTGGGCCACGGTCCGATTCGTATGGTTTTG GCGATTCTGGCGTTTTTTGCGTTTTTACCGCGATTAAACCGTCCCTGGGCC TGATTAATCGCTGGGGTTCCGTGGGCAAGAAAGAAGCGATGGAGATT ATCAAGAAATTTAAAAAAGATCTGGCAGCGATGCTTCGTATTATCAAT GCACGCAAAGAGCGTAAACGCCGCGGCGCGGATACGTCGATTGGCAT CATCGGTCTGCTGCTGACGACTGCAATGGCC
NS3hel	<sup>2</sup> atgGGTAAAACGGTGTGGTTTGTGCCGTCCGTGCGTAATGGCAACGAA ATTGCCGCCTGTCTTACCAAAGCAGGCAAACGCGTCATCCAGCTGAGT CGCAAAACCTTTGAAACCGAATTCCAGAAAACCAAAAACCAAGAATG GGATTTTCGTCATCACCACCGATATCAGCGAAATGGGCGCAAACCTTCAA AGCAGACCGCGTGATCGACAGCCGCCGCTGCCTGAAACCGGTCATCCT GGACGGCGAACGCGTCATCCTCGCCGGCCCAATGCCAGTTACCCATGC AAGCGCAGCGCAACGCCGTGGCCGCATCGGCCGTAACCCAAACAAGC CGGGTGACGAATATATGTAT

<sup>2</sup>Nota: O codão de iniciação (ATG), correspondente ao resíduo metionina, foi introduzido na sequência do *primer F*.

## ANEXO II

Tabela Anexo II. Informação relativa à composição dos meios, soluções e géis utilizados.

Meio/Solução/Gel	Componentes
<b>Tampão TAE 50X</b>	242g Tris-base 57,1ml ácido acético glacial 100ml 0,5 M EDTA, pH 8,0 Adicionar água para 1000 ml
<b>Gel de agarose a 1,5% (m/v) (50ml)</b>	0,75g agarose 50ml TAE 0,5X
<b>Gel de agarose a 1% (m/v) (50ml)</b>	0,5g agarose 50ml TAE 0,5X
<b>Gel de agarose a 0,8% (m/v) (50ml)</b>	0,4g agarose 50ml TAE 0,5X
<b>Meio de cultura LB</b>	LB (MB14502 - NZYtech, Portugal) *15g agar (MB02902 - NZYtech, Portugal) Adicionar água para 1000ml e autoclavar * Não adicionar para meio líquido
<b>Solução TSS</b>	LB 1X 10% (v/v) PEG6000 5% (v/v) DMSO 50mM MgSO <sub>4</sub> (pH 6,5) Filtrar
<b>Tampão de amostra (SDS-PAGE)</b>	0,5M Tris-HCl, pH 7,4 20% (v/v) glicerol 10% (m/v) SDS 5% (m/v) β-mercaptoetanol 0,5% (m/v) azul de bromofenol
<b>Gel de concentração a 4% (m/v) (tricina-SDS-PAGE) (10ml)</b>	1,32ml acrilamida/bisacrilamida (29:1) 30X 1,52ml 2,5M Tris-HCl, pH 8,8 100μl 10% (m/v) APS 10μl TEMED Adicionar água para 10ml
<b>Gel de resolução a 12% (m/v) (tricina-SDS-PAGE) (20ml)</b>	8ml acrilamida/bisacrilamida (29:1) 30X 11,2ml 2,5M Tris-HCl, pH 8,8 180μl 10% (m/v) APS 15μl TEMED Adicionar água para 20ml
<b>Gel de concentração a 5% (m/v) (SDS-PAGE) (10ml)</b>	1,69ml acrilamida/bisacrilamida (29:1) 30X 1ml 1,25M Tris-HCl, pH 6,8 100μl 10% (m/v) SDS 100μl 10% (m/v) APS 10μl TEMED Adicionar água para 10ml
<b>Gel de resolução a 13% (m/v) (SDS-PAGE) (20ml)</b>	8,9ml acrilamida/bisacrilamida (29:1) 30X 3ml 2,5M Tris-HCl, pH 8,8 200μl 10% (m/v) SDS 160μl 10% (m/v) APS 16μl TEMED Adicionar água para 20ml
<b>Tampão de lise (<i>refolding</i> proteico)</b>	20mM Tris-HCl, pH 7,4 0,5M NaCl 10% (v/v) glicerol 0,1% (v/v) Triton X-100 8M ureia

<b>Tampão de lise (purificação em condições desnaturantes)</b>	10mM Tris-HCl, pH 8 100mM NaH <sub>2</sub> PO <sub>4</sub> 7M ureia
<b>Tampão de lavagem (purificação em condições desnaturantes)</b>	10mM Tris-HCl, pH 6,3 100mM NaH <sub>2</sub> PO <sub>4</sub> 8M ureia
<b>Tampão de eluição (purificação em condições desnaturantes)</b>	10mM Tris-HCl, pH 4,5 100mM NaH <sub>2</sub> PO <sub>4</sub> 8M ureia
<b>Tampão de lise (purificação em condições nativas)</b>	50mM tampão fosfato de sódio, pH 7,1 0,5M NaCl 20mM imidazole, pH 7,4 10% (v/v) glicerol 0,1% (v/v) Triton X-100
<b>Tampão de lavagem (purificação em condições nativas)</b>	50mM tampão fosfato de sódio, pH 7,1 0,5M NaCl 30mM imidazole, pH 7,4 10% (v/v) glicerol 0,1% (v/v) Triton X-100
<b>Tampão de eluição (purificação em condições nativas)</b>	50mM tampão fosfato de sódio, pH 7,1 0,5M NaCl 0,5M imidazole, pH 7,4 10% (v/v) glicerol 0,1% (v/v) Triton X-100
<b>Tampão Towbin</b>	25mM Tris-base 192mM glicina 20% (v/v) metanol 0,5% (m/v) SDS
<b>Tampão TBS</b>	20mM Tris-HCl, pH 7,4 140mM NaCl
<b>Tampão TBS-T</b>	0,1% (v/v) Tween 20 em TBS
<b>Tampão de bloqueio</b>	2% (m/v) <i>ECL<sup>TM</sup> Advance Blocking Agent</i> (GE Healthcare Life Sciences, E.U.A.) em TBS-T
<b>Solução de detecção</b>	razão <i>ECL Advance Solution A/ ECL Advance Solution B</i> de 1:1 (GE Healthcare Life Sciences, E.U.A.)